UNIVERSITY OF LIVERPOOL

Identification and characterisation of novel vaccine targets for the control of trypanosomiasis

Thesis submitted in accordance with the requirements of the University of Liverpool for the degree of Doctor in Philosophy by:

Aitor Casas-Sanchez

June 2018

Liverpool School of Tropical Medicine
University of Liverpool
Liverpool, United Kingdom

ABSTRACT

Identification and characterisation of novel vaccine targets for the control of trypanosomiasis

Trypanosoma brucei is a unicellular protozoan parasite that causes human sleeping sickness and animal trypanosomiasis in sub-Saharan Africa. The parasite alternates its life cycle between a mammalian host and an insect vector (*Glossina* spp.). Within tsetse flies, trypanosomes undergo a series of developmental changes, which allow the parasite to adapt to the different tsetse tissues. Remodelling of the parasite surface glycocalyx is completed when the abundant variant surface glycoproteins (VSG) are expressed by the infectious metacyclic trypomastigote form (MCF) in the tsetse salivary glands. Although the role of the metacyclic VSG coat remains unknown, it is assumed this coat is important for parasite transmission and that it may mask or prevent the surface exposure of invariant proteins, in analogy with the blood stages. However, investigating metacyclic biology has been hindered as MCFs have been impossible to culture *in vitro* and harvesting enough parasites from infected tsetse salivary glands to perform biochemical analyses is challenging. In this thesis, I used a combination of several biochemical, genetic and cell biology methodologies to search for suitable transmission-blocking vaccine candidates to help control trypanosomiasis caused by *T. brucei* parasites.

A proteomics approach was used to identify parasite soluble factors found in *T. brucei*-infected tsetse saliva that may be crucial for transmission or early infection stages of the host. I confirmed that infected saliva is enriched with *T. brucei* surface proteins, including VSGs, Brucei Alanine-Rich Proteins (BARP) and a small, novel family of five GPI-anchored glycoproteins presently designated Metacyclic Invariant Surface Proteins (MISP). Notably, *misp* genes are highly transcribed in salivary gland-associated trypanosomes, but are predominantly expressed in the metacyclic stage. The crystal structure of the MISP360 isoform, together with molecular modelling, suggests that the MISP N-terminal ectodomains adopt a three-helical bundle conformation that is projected above the metacyclic VSG coat, thus exposing immunogenic epitopes. When *misp* genes were silenced by RNAi, a tsetse infection phenotype was not observed although these proteins may be important for parasite transmission rather than infection establishment. Collectively, these data suggest the *T. brucei* metacyclic trypomastigote surface is not just densely covered with a VSG coat, but rather it is more variable and allows for the co-expression of invariant GPI-anchored glycoproteins, like BARP and MISP, the latter which is more abundant and completely accessible to the hosts' antibodies. Due to these unique characteristics, MISP are predicted to be a strong vaccine candidate against trypanosomiasis.

In order to perform further functional studies on MISP, a system to generate *T. brucei* MCFs *in vitro*, based on the overexpression of trypanosomal RNA-binding protein 6 (RBP6), was developed. In a previous transcriptomic study, designed to characterise *T. brucei* genes involved in parasite migration to the tsetse salivary glands and/or metacyclogenesis, RBP6 was also identified among the top hits. Top candidates included a gene encoding for a hypothetical protein with an identifiable MYND-like zinc finger domain, three different folate transporter genes, and glutamate dehydrogenase. These genes were characterised *in vitro* and *in vivo* after they were expressed in a mutant parasite strain that fails to establish a salivary gland infection.

Strikingly, over expression of *rbp6* did not rescue the infection phenotype but the gene encoding for a MYND-like protein restored salivary gland infectivity and metacyclic formation. The *T. brucei* MYND-like protein function is unknown, but the presence of a zinc finger domain, together with its cytosolic localisation, suggest it may be involved in the regulation of parasite gene expression.

Since most of the *T. brucei* surface proteins are known to be *N*-glycosylated, I conducted a series of preliminary experiments to address the *in vivo* essentiality of parasite *N*-glycosylation in the tsetse. First, among all oligosaccharyltransferase genes (*sst3a-c*), *sst3b* is the main isoform expressed during *T. brucei* development in the tsetse. Given its glycan specificity, this suggests that oligomannose *N*-glycans are likely the predominant oligosaccharide in the insect stages. Second, functional studies suggest that protein *N*-glycosylation may be essential for parasite establishment in the tsetse salivary glands. Collectively, these results suggest that the *T. brucei N*-glycosylation pathway could be exploited to generate new transmission-blocking approaches. Lastly, I developed a new methodology for imaging tsetse tissues in great detail. We can now generate complete 3D reconstructions of tissues using laser scanning confocal microscopy, to study, for instance, the interactions between trypanosomes and tsetse or determine changes in the architecture of tsetse tissues with different stressors.

In summary, this thesis focused on the *T. brucei* stages that colonise the tsetse proventriculus and salivary glands to identify key molecules involved with parasite differentiation and transmission. This work carries potential implications for transmission-blocking approaches against African trypanosomiases, mainly (but not exclusively) caused by *T. brucei*. The discoveries of MISP, the hypothetical zinc finger protein MYND and the essentiality of *N*-glycosylation have revealed important targets for development in the tsetse, which can be studied at a higher level using our new imaging methods.

DECLARATION

I hereby certify that this dissertation constitutes my own product, that where the language of others is set forth, quotation marks so indicate, and that appropriate credit is given where I have used the language, ideas, expressions or writings of another.

I declare that the dissertation describes original work that has not previously been presented for the award of any other degree of any institution.

Signed,

……

TABLE OF CONTENTS

ABSTRACT **2**

DECLARATION **4**

TABLE OF CONTENTS **5**

LIST OF FIGURES **10**

LIST OF TABLES **13**

ABBREVIATIONS **14**

ACKNOWLEDGEMENTS **16**

CHAPTER 1 **17**

1.1 General Introduction 18

1.2 Human African Trypanosomiasis 18

1.2.1 Importance of the disease 18

1.2.2 Epidemiology 19

1.2.3 Clinical presentations 20

1.2.4 Diagnosis 21

1.2.5 Treatment 22

1.2.6 Disease control 23

1.3 *Trypanosoma brucei* biology 24

1.3.1 Cell biology 24

1.3.2 Life cycle 25

1.4 *Glossina* biology 28

1.5 *Trypanosoma brucei* major surface glycoproteins 30

1.5.1 Variant Surface Glycoproteins and antigenic variation 30

1.5.2 Metacyclic Variant Surface Glycoproteins 33

1.5.3 Procyclins 34

1.5.4 Brucei Alanine-Rich Proteins (BARPs) 34

1.6 Gene expression regulation in *Trypanosoma brucei* 35

1.7 *N*-glycosylation 37

1.7.1 *N*-glycan structures 37

1.7.2 *N*-glycan biosynthesis and processing 38

1.7.3 *N*-glycosylation in *Trypanosoma brucei* 40

1.7.4 Glycosylphosphatidylinositol (GPI) anchors 41

1.7.5 *N*-glycan and GPI anchor structures in *Trypanosoma brucei* 42
1.7.6 Functional studies on *Trypanosoma brucei N*-glycosylation 44
1.8 Aims of the thesis 46

CHAPTER 2 **47**
2.1 Cell culture methods 48
2.1.1 *Trypanosoma b. brucei* bloodstream forms 48
2.1.2 *Trypanosoma b. brucei* procyclic cultured forms 48
2.1.3 *Trypanosoma b. brucei* stabilates 48
2.1.4 Induction of the tetracycline-inducible genetic system in *T. b. brucei* 49
2.1.5 *Escherichia coli* culture 51
2.1.8 *Escherichia coli* stocks 52
2.2 Molecular biology methods 52
2.2.1. Extraction of *Trypanosoma brucei g*enomic DNA (gDNA) 52
2.2.2 Plasmid purification 53
2.2.3 Non-genomic DNA purification 53
2.2.4 RNA extraction and cDNA synthesis 53
2.2.5 Polymerase chain reaction (PCR) 54
2.2.6 Primer design 55
2.2.7 Agarose gel electrophoresis 55
2.2.8 Semi-quantitative RT-PCR 56
2.2.9 Real-time quantitative RT-PCR (qRT-PCR) 56
2.2.10. Restriction of DNA 57
2.2.11 Plasmid dephosphorylation 57
2.2.12. DNA ligation 57
2.2.13. *Escherichia coli* transformation 58
2.2.14. Plasmid linearisation 58
2.2.15. DNA precipitation 59
2.2.16 General cloning procedure 59
2.2.17. Creation of plasmids for the expression of MISP recombinant proteins 60
2.2.18 Creation of plasmids for the generation of *T. brucei* mutant lines 61
2.3 Methods for protein analysis 67
2.3.1 Expression of recombinant proteins in *E. coli* 67
2.3.2 Ni-NTA chromatography 68

2.3.3 Size exclusion chromatography (SEC) 68
2.3.4 Production of anti-MISP polyclonal antibodies 69
2.3.5. SDS-PAGE 69
2.3.6 Western blotting 69
2.3.7 Lectin blotting 71
2.3.8 Procyclin extraction from *T. brucei* PCF cells 71
2.3.9 nLC-MS/MS Proteomics 72
2.3.10 Protein pulldown assay 73
2.4 Tsetse methods 73
2.4.1 Tsetse fly infections and maintenance 73
2.4.2 Tsetse dissections and *T. brucei* collection 74
2.5 Cell biology methods 74
2.5.1 Immunofluorescence assay 74
2.5.2 Live immunostaining 76
2.5.3 Laser Scanning Confocal Microscopy (LSCM) 76
2.5.3 Fluorescence-Activated Cell Sorting (FACS) 76
2.6 Ion-exchange chromatography 77
2.7 Bioinformatics 77
2.8 Statistics 78

CHAPTER 3 **79**
3.1 Introduction 80
3.2 Results 81
3.2.1 Mass spectrometry of trypanosome-infected tsetse saliva 81
3.2.2 *In silico* analyses on MISP 91
3.2.3 Production of MISP recombinant proteins 100
3.2.4 Creation of a polyclonal antibody against MISP 105
3.2.5 Gene and protein expression analyses on MISP and BARP 109
3.2.6 Structural studies on MISP and BARP 121
3.2.7 Functional studies on MISP 132
3.3 Discussion 136
3.3.1 *T. brucei*-infected tsetse saliva is enriched with trypanosome GPI-anchored surface glycoproteins. 136
3.3.2 Do metacyclic VSGs recombine? 137
3.3.3 The Metacyclic Invariant Surface Proteins (MISP) 138

3.3.4 Production of MISP recombinant proteins for the study of MISP 139
3.3.5. Expression studies on MISP and BARP 140
3.3.6 Insights into the protein structures of MISP and BARP 141
3.3.7 The metacyclic surface glycocalyx 142
3.3.8 Function of MISP 143
3.3.9 Uses of MISP and future work 144

CHAPTER 4 **146**
4.1 Introduction 147
4.2 Results 148
4.2.1 Creation of novel *T. brucei* cell lines overexpressing RBP6 148
4.2.2 Differentiation dynamics of RBP6 overexpressor cells 150
4.2.3 Ion-exchange purification of *in vitro* MCF and BSF-like cells 153
4.2.4 Developmental expression of major surface glycoproteins 154
4.2.5 Developmental expression of RBP6 157
4.2.6 *In vivo* phenotype of RBP6 overexpression in the tsetse 158
4.2.7 Assessing the essentiality of RBP6 for parasite development in the tsetse 159
4.2.8 Identification of other *T. brucei* genes important to progress towards a SG infection and transmission 161
4.3 Discussion 176
4.3.1 Overexpression of trypanosomal RBP6: a tool to study *T. brucei* development *in vitro* 176
4.3.2 Gene expression potentially important for *T. brucei* transmission 178

CHAPTER 5 **180**
5.1 Introduction 181
5.2 Results 182
5.2.1 Developmental expression of *stt3* paralogs 182
5.2.2 Functional studies on *N*-glycosylation 184
5.2.3 Systematic collection of *T. brucei* metacyclic cells for glycan structural analysis 191
5.3 Discussion 193
5.3.1 Developmental regulation of the trypanosomal STT3 oligosaccharyltransferase isoforms 193

5.3.2 Functional studies on *N*-glycosylation in the tsetse stages 194

CHAPTER 6 **197**

6.1 Introduction 198

6.2 Methods 198

6.2.1 Tsetse nitisinone feeding and tissue processing 199

6.2.2 Live staining of the tsetse peritrophic matrix 199

6.2.3 3D reconstruction of trypanosome-infected tsetse proventriculus 200

6.2.4 *Ex vivo* 3D reconstruction of trypanosome-infected peritrophic matrices 200

6.3 Results 201

6.3.1 3D reconstruction of tsetse organs to investigate tissue damage 201

6.3.2. Studies on *T. brucei*-tsetse interactions using imaging methods 208

6.4. Discussion 216

6.4.1 New imaging methodology to reconstruct tsetse tissues 216

6.4.2. Assessing tsetse tissue damage upon nitisinone exposure 217

6.4.3. Studies on a novel alternative route of tsetse PV infection by *Trypanosoma brucei* 219

CHAPTER 7 **221**

REFERENCES **227**

APPENDIX I **250**

LIST OF FIGURES

Chapter 1. Introduction

Figure 1.1 Geographical distribution of HAT and AAT trypanosomes 20
Figure 1.2 *T. brucei* cell architecture 25
Figure 1.3 *T. brucei* life cycle 27
Figure 1.4 Variant Surface Glycoproteins and antigenic variation 32
Figure 1.5 Gene expression regulation in *T. brucei* 36
Figure 1.6 *N*-glycan structural types 38
Figure 1.7 *N*-glycosylation synthesis pathway 40
Figure 1.8 *N*-glycan and GPI glycan structures in *T. brucei* 43

Chapter 2. Materials and methods

Figure 2.1 Transgene insertions for the tetracycline-inducible system in *T. b. brucei* parental lines AnTat 1.1 90:13 and Lister 427 29:13 49
Figure 2.2 The tetracycline-inducible genetic system in *T. brucei* 50
Figure 2.3 Calculations on insert amounts in a DNA ligation reaction 58

Chapter 3. The Metacyclic Invariant Surface Proteins (MISP)

Figure 3.1 Workflow for the analysis of *T. brucei*-infected tsetse saliva 82
Figure 3.2 GO terms distribution *T. brucei* proteins identified in tsetse saliva 83
Figure 3.3 Multiple protein alignment of *T. brucei* BARPs 87
Figure 3.4 Mass spectra of BARP and mVSG detected by nLC-MS/MS 88
Figure 3.5 Mass spectrum of MISP detected by nLC-MS/MS 90
Figure 3.6 Immunoblotting of tsetse saliva probing with anti-Sodalis mAb 91
Figure 3.7 Schematic of the genomic arrangement of *misp* homologs 93
Figure 3.8 Phylogenetics and multiple protein alignment of MISP 94
Figure 3.9 Multiple protein alignment of all MISP isoforms 95
Figure 3.10 PROSPER protease cleavage site predictions on MISP 98
Figure 3.11 3D structural models of MISP360 and MISP400 99
Figure 3.12 Expression and Ni-NTA purification of MISP380 101
Figure 3.13 SEC purification of MISP380 102
Figure 3.14 Engineered DNA constructs for the expression of recombinant variations of MISP360 in *E. coli* 104
Figure 3.15 Ni-NTA purification of recombinant MISP360 variations 106
Figure 3.16 Schematic of the workflow for the creation of a polyclonal α-MISP 107

Figure 3.17 Reactivity tests of α-MISP polyclonal antibody 108
Figure 3.18 Exploiting the MISP C-termini repeats to determine expression 110
Figure 3.19 RNA expression of *misp* and *barp* throughout the *T. brucei* life cycle 112
Figure 3.20 Controls and validation of the RNA expression analyses 113
Figure 3.21 Immunostaining of *T. brucei* infecting the tsetse salivary glands using α-MISP and α-BARP 115
Figure 3.22 Expression and localization of MISP and BARP in SG stages 117
Figure 3.23 Live immunostaining of MCF with α-MISP 118
Figure 3.24 Ectopic localization of MISP360 119
Figure 3.25 Indirect localization of mVSG in SG stages 121
Figure 3.26 The crystal structure of MISP360 N-terminus 122
Figure 3.27 High confidence model of BARP 124
Figure 3.28 Comparison between MISP, BARP and VSG C-termini 125
Figure 3.29 Modelling the five MISP isoforms on the MCF surface 126
Figure 3.30 MISP structural homologs 127
Figure 3.31 Molecular pockets on the surface of MISP and BARP 129
Figure 3.32 Structural model of the MCF surface glycocalyx 130
Figure 3.33 Developmental expression of major surface glycoproteins during the life cycle of *T. brucei* 131
Figure 3.34 Essentiality of MISP for tsetse infection 133
Figure 3.35 Pulldown assays on MISP400 and tsetse saliva 135

Chapter 4. Generation of *in vitro* MCF for the study of MISP

Figure 4.1 Overexpression of RBP6 in *T. brucei* PCF 149
Figure 4.2 Ectopic localization of RBP6 150
Figure 4.3 RBP6-induced differentiation *in vitro* 152
Figure 4.4 *T. brucei* life stages generated *in vitro* during RBP6 overexpression 153
Figure 4.5 Detection of developmentally regulated surface glycoproteins in RBP6 overexpressor cells 156
Figure 4.6 Developmental expression of RBP6 throughout the life cycle of *T. brucei* 157
Figure 4.7 *In vivo* phenotype of RBP6 overexpressors in the tsetse 159
Figure 4.8 Essentiality of RBP6 160
Figure 4.9 Comparison of wildtype *T. b. brucei* Lister 427 and TSW-196 162

Figure 4.10 Differential expression gene analysis between strains Lister 427 and TSW-196 proventricular trypanosomes 164
Figure 4.11 3D models of the candidate proteins 167
Figure 4.12 Gene expression of the candidate genes during life cycle 168
Figure 4.13 Gene expression of the hypothetical MYND-like protein 169
Figure 4.14 Gene expression of the folate transporters 170
Figure 4.15 Gene expression of the glutamate dehydrogenase 171
Figure 4.16 Gene expression of RBP6 172
Figure 4.17 DNA constructs for the overexpression of the candidate genes 173
Figure 4.18 Cellular localization of ectopically-expressed candidate proteins 174
Figure 4.19 *In vivo* phenotype of candidate genes overexpression 175

Chapter 5. Studies on *N*-glycosylation in the tsetse stages of *T. brucei*
Figure 5.1 Developmental expression of the oligosaccharyltransferase paralogs 183
Figure 5.2 DNA constructs for the creation of conditional-null STT3 and RFT1 mutant trypanosomes 185
Figure 5.3 Genotyping of conditional-null STT3 and RFT1 mutants 186
Figure 5.4 Characterisation of *stt3* RNAi cell line 188
Figure 5.5 Characterisation of procyclins from *stt3* RNAi cells 190
Figure 5.6 Essentiality of *N*-glycosylation for tsetse infection 191
Figure 5.7 SG infection rates obtained for the collection of MCFs 192

Chapter 6. New methods for imaging African trypanosomes in the tsetse
Figure 6.1 Tsetse tissues investigated for damage upon nitisinone exposure 202
Figure 6.2 Midgut and proventriculus reconstructions 203
Figure 6.3 Reconstructions of tsetse flight muscle 205
Figure 6.4 Tsetse fat bodies reconstructions 206
Figure 6.5 Malpighian tubules reconstructions 207
Figure 6.6 Tsetse salivary gland reconstructions 207
Figure 6.7 *T. brucei* early infection of the tsetse proventriculus 209
Figure 6.8 *T. brucei* localization in the proventriculus and anterior MG 211
Figure 6.9 3D reconstruction of infected proventriculus 212
Figure 6.10 *T. brucei* pockets in the tsetse peritrophic matrix 214
Figure 6.11 Single *T. brucei* cells interacting with the tsetse PM 215

LIST OF TABLES

Table 2.1 Antibiotics used for the selection of transformant *E. coli* 52
Table 2.2 Plasmids used for the creation of *T. brucei* transgenics 65
Table 2.3 Drugs used for the selection of transfectant *T. brucei* 67
Table 2.4 Antibodies used in western blotting 70
Table 2.5 Antibodies used in immunofluorescence assay 75
Table 3.1 Intracellular *T. brucei* proteins identified by nLC-MS/MS in infected tsetse saliva 84
Table 3.2 *T. brucei* GPI-anchored surface proteins detected in infected tsetse saliva by nLC-MS/MS 85
Table 3.3 Summary of *T. brucei* VSG species found in either cultured or tsetse-derived metacyclics (literature) 89
Table 3.4 Summary of *in silico* predictions on MISP protein sequences 97
Table 4.1 Top up-regulated *T. b. brucei* genes in TSW-196 compared to Lister 427 29: 13 proventricular trypanosomes 165

ABBREVIATIONS

AAT	Animal African trypanosomiasis
AM	Anterior midgut
BES	Bloodstream Expression Site
BSF	Bloodstream form
cBSF	Cultured BSF
cDNA	Complementary DNA
CNS	Central nervous system
CRD	Cross-reacting determinant
CTD	C-terminal domain
DIC	Differential interference contrast
DNA	Deoxyribonucleic acid
Dox	Doxycycline
dpi	Days post infection/induction
dsRNA	Double stranded RNA
EC	Endothelial cell
EMF	Epimastigote form
ER	Endoplasmic reticulum
ES	Ectoperitrophic space
ESAG	Expression Site Associated Gene
FBS	Faetal bovine serum
FT	Folate transporter
GDH	Glutamate dehydrogenase
gDNA	Genomic DNA
GFP	Green fluorescent protein
GPI	Glycosylphosphatidylinositol
HAT	Human African trypanosomiasis
IgG	Immunoglobulin G
kDNA	Kinetoplastid / mitochondrial DNA
KO	Knockout
LB	Lysogenic broth
MCF	Metacyclic form
MES	Metacyclic Expression Site

MIP	Maximum intensity projection
MG	Midgut
MSC	Mesocyclic form
mVAT	Metacyclic Variant Antigen Type
mVSG	Metacyclic VSG
MYND	MYND-like protein
NFW	Nuclease-free water
NTBC	Nitisinone
OST	Oligosaccharyltransferase
P-MCF	Pre-metacyclic form
PBS	Phosphate buffered saline
PCF	Procyclic cultured form
PCR	Polymerase chain reaction
PF	Procyclic form
PFA	Paraformaldehyde
PM	Peritrophic matrix
PV	Proventriculus
qPCR	Quantitative real-time PCR
RBP	RNA-binding protein
RNA	Ribonucleic acid
RNA Pol	RNA polymerase
RNAi	RNA interference
RT	Room temperature
RT-PCR	Reverse transcription PCR
SDS-PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
SG	Salivary gland
SL	Slender
ST	Stumpy
Tet	Tetracycline
UTR	Untranslated region
VSG	Variant Surface Glycoprotein
WHO	World Health Organization
3D	3-dimensions

ACKNOWLEDGEMENTS

First, I'd like to thank my supervisor Dr. Álvaro Acosta-Serrano for giving me, in first place, the opportunity to carry out this thesis which has allowed me to do what I enjoy the most, and it has changed my life forever. Thank you for your unconditional support and for being more than a supervisor in all aspects. I would also like to thank my former supervisors Dr. Cristina Madrid, Dr. Sonia Paytubi, Dr. Valerie Gerard and Prof. Iouri Gounko for the given opportunities and support which led me to this moment.

A huge especial thanks to my parents, Alex and Ana, my grandparents Rafa, Ana, Joni and Cinta, and the rest of uncles, aunts and cousins who have always given me the best they could, and without their support and love I wouldn't be writing these lines. A special mention to Carla who has closely witnessed the entire path with patience, generosity and love.

I can't forget about many old friends who, despite the distance, are still there for me: Uri, Guiu, Victor, Álvaro, Alex, Oriol, Carmen, Irene, Elena *et al*. Also, new good friends I made during my stay in Liverpool, who have completely made this experience more enjoyable: Danilo, Karina, Nelson, Cintia, Jacob, Aldo, Silvia, Cris, Selene, Kris, Rocio, *et al.*

A special mention to the GlycoPar people and to the glycobiology field that brought us together, and to the whole BoP community for an unforgettable once in a lifetime experience.

A big thank you to all my lab and LSTM colleagues and friends who have somehow contributed to the success of this thesis (or not): Lee, Rob, Karina, Kris, Clair, Yasser, Raquel, Vicki, Keith, James, Dan, Samïrah, Naomi, Yuk-Chien, Nich, Waleed, Jesús, Elena, Eliza, Nadia, Dee, *et al.*

Also, massive thanks to the students I supervised for their contribution to this work: Sarah, Lewis, Rob, Cintia, Laura, Lara, Carla, Zena and Barrack.

Love you all

CHAPTER 1

Introduction

1.1 General Introduction

Neglected tropical diseases (NTDs) are a group of 20 human infections (WHO, 2018) affecting over a billion people, with major incidence in tropical and sub-tropical countries across the globe. Most NTDs are vector-borne and are caused by a broad range of infectious pathogens such as bacteria, viruses, protozoa and helminths. The burden of NTDs especially affects human populations from low-income countries with poor sanitation and in proximity with the disease vectors and livestock. Disease control relies on a combination of public health approaches such as vector control, screening of populations and mass drug administration. In summary, most NTDs, including those caused by kinetoplastid parasites like Leishmaniasis and African trypanosomiasis, have a huge impact in the people, economy and development of endemic countries (Feasey et al., 2010).

1.2 Human African Trypanosomiasis

1.2.1 Importance of the disease

African trypanosomiasis is a group of vector-borne diseases caused by protozoan parasites and mainly transmitted by tsetse flies, which affect both human and livestock in sub-Saharan African countries where the vector is found (Barrett et al., 2003; Malvy and Chappuis, 2011). Human African trypanosomiasis (HAT) cases have declined every year since 2010 due largely to a program of active screening and treatment of the human population in endemic areas, together with an efficient vector control program (2,184 cases reported in 2016, WHO). However, it is thought that these numbers are a vast underestimation since many cases occur in rural areas with no access to healthcare (Fevre et al., 2008; Simarro et al., 2012). Animal African trypanosomiasis (AAT or *Nagana* disease), however, continues to cause huge economic losses in the agricultural sector with more than 1 million cattle dying of AAT each year (Giordani et al., 2016). Despite the low number of reported HAT cases, over 70 million people continue to be at risk and so more efforts need to be made towards its elimination, for which the World Health Organization (WHO) has overambitiously set for the year 2020.

1.2.2 Epidemiology

There are currently several parasitic diseases in sub-Saharan Africa known as African trypanosomiasis, which affect both humans and animals. HAT, commonly known as sleeping sickness, can be caused by two morphologically identical subspecies of the unicellular parasite *Trypanosoma brucei*; i.e. *T. brucei gambiense* and *T. brucei rhodesiense* (Simarro et al., 2010). These subspecies have different geographic distribution (**Figure 1.1**) and present distinct clinical manifestations. *T. b. gambiense* is found in Central and West Africa, causing a slow chronic form of the disease, in where humans are thought to be the main reservoirs of this anthroponotic disease. *T. b. gambiense* HAT was responsible for the ~95% of cases in 2016 (WHO), 75% of them reported in the Democratic Republic of the Congo. On the other hand, *T. b. rhodesiense* prevails in South and East Africa and causes an acute and virulent form of the disease which is mainly zoonotic with multiple animal reservoirs. Only in Uganda there is transmission of both subspecies (Picozzi et al., 2005) and, although human co-infections have not been reported, *T. b. rhodesiense* cattle reservoirs have been recently described to migrate towards *T. b. gambiense* endemic areas in North Uganda, which could potentially lead to co-infections. Cattle are particularly affected by AAT, which is mainly caused by *T. congolense*, *T. vivax* and *T. evansi*. AAT is one of the main causes of death for cattle in sub-Saharan Africa, burdening meat and milk production which translates in losses of billions of US dollars and halting development. AAT spreads beyond the sub-Saharan Africa (**Figure 1.1**) since species like *T. vivax* or *T. evansi*, which do not have a complete life cycle in the tsetse, can be mechanically transmitted by non-tsetse insect vectors such as horseflies (*Tabanus* spp.) or stable flies (*Stomoxys* spp.) (Giordani et al., 2016).

HAT is primarily transmitted to humans through the bite of the tsetse fly biological vectors, although transmission may eventually occur through blood transfusion (Herwaldt, 2001; Hira and Husein, 1979), congenital or even sexual routes (De Kyvon et al., 2016; Lindner and Priotto, 2010; Rocha et al., 2004). All 31 *Glossina* species, classified as forest, savannah and riverine types according to habitat, have the potential to transmit African trypanosomes and their geographical distribution defines the distribution of HAT (Cecchi et al., 2008; Jordan, 1993). The most significant tsetse species responsible for *T. b. gambiense* HAT transmission are the riverine *G. palpalis* and *G. fuscipes* (Grebaut et al., 2016), whereas the savannah

species *G. morsitans*, *G. pallidipes* and riverine *G. f. fuscipes* are the main responsible for transmitting *T. b. rhodesiense* HAT (Shereni et al., 2016).

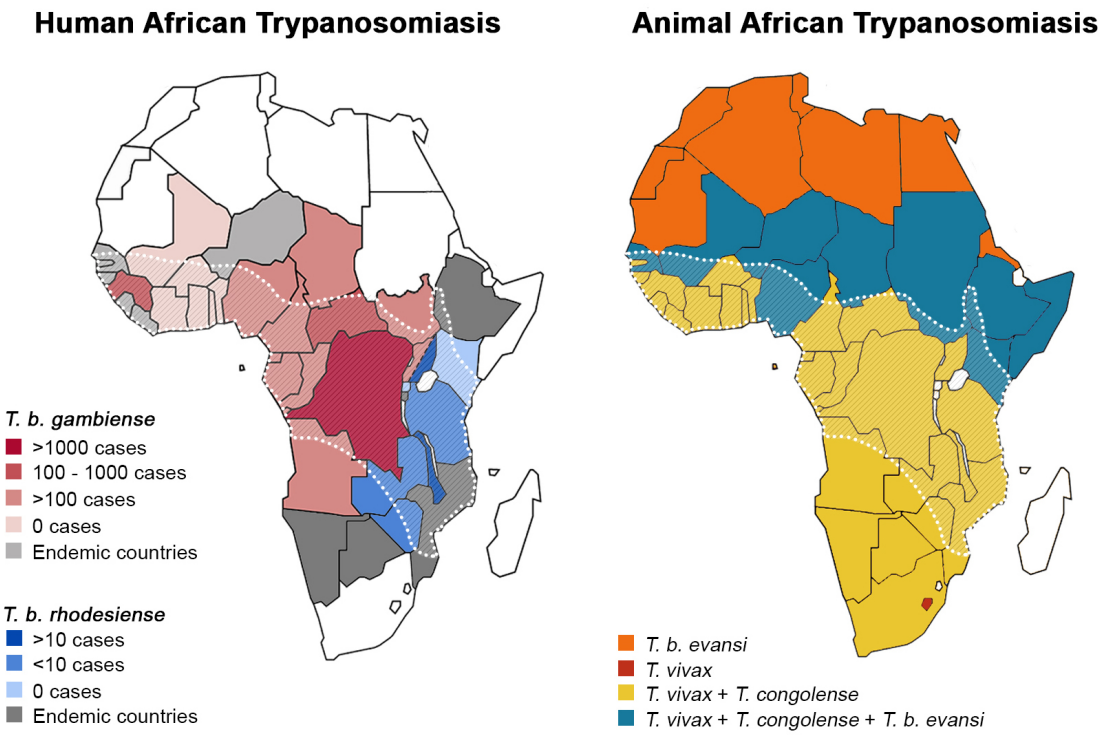


Figure 1.1. Geographical distribution of trypanosome species causing human African trypanosomiasis and animal African trypanosomiasis in the African continent. Distribution of *T. b. gambiense* and *T. b. rhodesiense* based on the number of reported cases in 2016 (left). Distribution of *T. evansi*, *T. vivax* and *T. congolense*. The striped white-dotted area represents the habitat distribution of Glossina spp. Modified from WHO (HAT) and Giordani et al. 2016 (AAT).

1.2.3 Clinical presentations

The clinical presentations for sleeping sickness vary depending on the parasite subspecies, stage of the disease and the response of the host. Both forms of the disease are usually fatal if they remain untreated, although individuals may

eventually become asymptomatic or self-heal (Jamonneau et al., 2000; Jamonneau et al., 2002; Jamonneau et al., 2012). The diseases typically go through two clinical stages. In the haemolymphatic stage (stage 1) the parasites remain in the blood and lymph, and in some cases, they can invade the adipose tissue and the skin (Trindade et al., 2016; Capewell et al., 2016;). Symptoms include intermittent fever, headache and skin rash among others (Blum et al., 2006; Duggan and Hutchinson, 1966). To progress to the meningoencephalitic stage (stage 2), the parasites cross the blood-brain barrier invading the central nervous system (CNS) (Masocha et al., 2007) leading to neurological alterations that may include sleep disorders, behavioural changes, coma and death (Buguet et al., 2001). Recent studies on mice suggest that these sleeping disturbances (from which the disease is named) are a result of the manipulation of the host's circadian rhythm by the parasite (Rijo-Ferreira et al., 2018; Rijo-Ferreira et al., 2017a; Rijo-Ferreira et al., 2017b). While *T. b. gambiense* HAT takes years to progress (Checchi et al., 2008a; Checchi et al., 2008b), the *T. b. rhodesiense* disease is typically acute and progresses to the stage 2 within weeks and death within 6 months (Odiit et al., 1997).

1.2.4 Diagnosis

Disease diagnosis requires specific tests since the clinical symptoms are usually unspecific. The diagnosis of *T. b. gambiense* HAT often starts with the card agglutination test for trypanosomiasis (CATT) (Chappuis et al., 2002; Magnus et al., 1978), which can be done on capillary blood or serum, and it is fast enough to allow large-scale screening of populations in endemic areas (Robays et al., 2004). However, due to its sub-optimal sensitivity (Jamonneau et al., 2015), parasitological validation by microscopic examination of capillary blood or serum is usually required (Lutumba et al., 2005a). Since parasitaemia levels can be below the detection threshold, concentration methods may be required such as the microhaematocrit or anion-exchange centrifugation techniques (Woo, 1970) (Mumba Ngoyi, 2014). In contrast, diagnosis of *T. b. rhodesiense* infections relies on the unspecific clinical manifestation and the history of exposure of the patient because there is no serological test available. Microscopy tests, however, are more reliable as the parasite densities in blood and lymph are usually high. The identification of stage 2 patients for both *T. b. gambiense* and *T. b. rhodesiense* diseases by microscopic

examination of the cerebrospinal fluid after lumbar puncture is of great importance as treatments differ between stages (Kennedy, 2008; Mumba Ngoyi et al., 2013).

1.2.5 Treatment

There is no preventive or therapeutic vaccine against African trypanosomiasis mainly because of antigenic variation, the phenomenon by which the parasite evades the host's immune system by periodically changing the variant surface glycoproteins (VSGs) coat (Horn, 2014). The only treatments for the diseases rely on the administration of five drugs in accordance to the parasite subspecies and stage of the infection. Pentamidine isethionate is used intramuscularly as a high-efficacy first-line treatment of stage 1 *T. b. gambiense* infections (Bronner et al., 1991). Intravenous infusion of suramin (Gustafsson, 1987), although effective against both subspecies, it is only used to treat stage 1 *T. b. rhodesiense* to prevent strong allergic reactions in patients co-infected with onchocerciasis in *T. b. gambiense*-endemic areas. To treat *T. b. gambiense* stage 2, a combination of nifurtimox (orally) and eflornithine (intravenously) therapy (NECT) is frequently used (Priotto et al., 2009). Lastly, intravenous melarsoprol is only used to treat stage 2 *T. b. rhodesiense* infections because of its high rate of severe side effects, including death (Schmid et al., 2005). The majority of these drugs have moderate rates of unpleasant side effects and low fatality rates, being melarsoprol the exception with the highest rates (Franco, 2012; Priotto et al., 2009; Schmid et al., 2012) . Remarkably, two new drugs (fexinidazole and SCYX-7158) that are currently in advanced phases of clinical trials can significantly improve these treatments (Jacobs et al., 2011; Tarral et al., 2014). They are orally administered in a few doses to treat both stages of *T. b. gambiense* and *T. b. rhodesiense* infections, removing the need for stage determination by lumbar puncture. Drug resistance has been described as a result of mutations in the parasite genome, especially against pentamidine and melarsoprol (Alsford et al., 2013; Graf et al., 2013; Munday et al., 2014). New combined therapies such as the NECT are in part designed to prevent drug resistance by using several drugs with different modes of action.

1.2.6 Disease control

HAT is mainly controlled due to a combination of active surveillance, screening and treatment methods, and vector control (Abel et al., 2004; Barrett, 2006; Lutumba et al., 2005b). Ceasing these activities may result in a re-emergence of the disease in endemic areas (Smith et al., 1998; Van Nieuwenhove et al., 2001). For instance, in the 1960s the number of sleeping sickness cases was similar to the current ones, but after control strategies halted there was resurgence (Abel et al., 2004; Lyons, 1992). The most effective approach for controlling *T. b. gambiense* HAT is case detection during big-scale screening and treatment, which aims to reduce the number of human reservoirs (Barrett, 2006). However, some studies suggest that domestic animals, not considered in most epidemiological studies, may be also reservoirs for *T. b. gambiense* (Kagbadouno et al., 2012; Njiokou et al., 2006; Simo et al., 2006). In addition, the recent discovery of the human skin (Capewell et al., 2016) and adipose tissue (Trindade et al., 2016) as parasite reservoirs, and asymptomatic carriers (Jamonneau et al., 2012) may be of great importance to improve the efficacy of current control methods. The control of *T. b. rhodesiense* HAT is mainly based on the control of domestic animals as the key disease reservoir (Enyaru et al., 2006; Welburn et al., 2001; Wendo, 2002), although wild animals also play a role in the epidemiology and are more difficult to tackle in control approaches. Additionally, vector control strategies which intend to reduce the incidence of tsetse in these foci are of great importance too. The use of insecticide-treated tsetse-attractive traps or “tiny targets” is the main vector control method currently used (Courtin et al., 2015), although aerial (and animal) insecticide spraying and sterile insect technique (or SIT) can also reduce tsetse populations depending on the geographical area and *Glossina* species (Shaw et al., 2013; Shaw et al., 2015; Torr et al., 2007).

WHO has appointed HAT for elimination as a public health problem (i.e. <1 annual case/10,000 people in a given focus) by 2020. However, eradication or total transmission interruption, set for 2030, seems to be a less achievable goal considering the current limitations in diagnosis and control of asymptomatic individuals or animal reservoirs, especially for *T. b. rhodesiense* HAT (Simarro et al., 2015).

1.3 *Trypanosoma brucei* biology

1.3.1 Cell biology

Trypanosoma brucei is a unicellular parasitic protozoa that belongs to the genus *Trypanosoma* and to the order kinetoplastida (Stevens, 2004). *T. brucei* is a diploid organism containing 11 chromosome pairs in the nucleus along with ~100 intermediate and mini-chromosomes, which mostly contain arrays of *vsg* genes (Berriman et al., 2005). The megabase-chromosomal genome size is of approximately ~ 26 Mb containing 9068 predicted genes, most of them lacking introns, and it goes up to ~35 Mb including smaller chromosomes and mitochondrial DNA. *T. brucei* has been extensively studied not only as the causative agent of sleeping sickness, but also because of its unique biology as primitive eukaryotic organism. Although the human pathogenic subspecies are *T. b. gambiense* and *T. b. rhodesiense*, the related animal-infectious *T. b. brucei* has been the main *T. brucei* subspecies used in research due to safety reasons and because of the high conservation between their genomes. *T. b. rhodesiense* is known to have an identical genome than *T. b. brucei* (Sistrom et al., 2016), except for the presence of the serum resistance antigen gene (Gibson, 2005; Zoll et al., 2018), which confers resistance to killing by human apolipoprotein A1 (APOL1), which is part of the trypanolytic factor (TLF) complex (Thomson et al., 2014; Vanhamme and Pays, 2004). Interestingly, *T. b. gambiense* lacks the *sra* gene, but its resistance to human serum relies on a slower endocytosis rate of the TFL (Uzureau et al., 2013). *T. brucei* is a flagellated organism that presents a broad variety of cell sizes and morphologies throughout its life cycle. As all eukaryote cells (**Figure 1.2**), it has one nucleus containing the chromosomal DNA, a complex reticular system and a mitochondrion whose morphology varies depending on the nutritional conditions. As in all kinetoplastids, *T. brucei* has a kinetoplast (kDNA) which contains the mitochondrial DNA formed by ~50 maxicircles and ~10,000 mini-circles, and it is part of the flagellum insertion complex at the flagellar pocket (Simpson and Simpson, 1980; Webster and Russell, 1993). Remarkably, the bloodstream form of *T. brucei* has modified peroxisomes called glycosomes in where glycolysis occurs in a compartmentalised manner (Opperdoes et al., 1984).

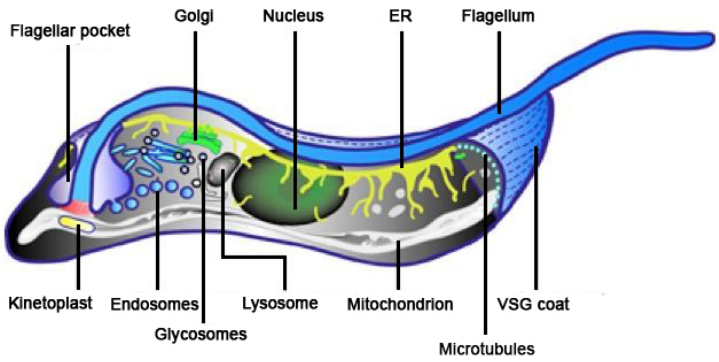


Figure 1.2. *T. brucei* cell architecture. Schematic of a *T. brucei* bloodstream form cell highlighting the major organelles. The DNA is contained in the nucleus and kinetoplast. The highly-developed endoplasmic reticulum (ER) and mitochondrion are found along the cell. The Golgi apparatus, lysosomes, endosomes and glycosomes (only in BSF cells) are usually found between the nucleus and the kinetoplast. The flagellar pocket is found at the flagellum insertion site in the cell body. The cell surface is usually coated by major surface glycoproteins (e.g. VSGs in BSF cells) and a complex grid of pellicular microtubules is found under the plasma membrane to control cell shape and motility (modified from Prof. Markus Engstler ®).

1.3.2 Life cycle

T. brucei is an obligate parasite that alternates its life cycle between a mammalian host and the tsetse fly vector (**Figure 1.3**). It continuously undergoes developmental changes to adapt and overcome the challenges found in the different environments in the hosts (Matthews, 2005). These include changes in cell morphology, motility, metabolism or glycocalyx composition among others, all ultimately a result of either induced or programmed changes in gene expression.

When a tsetse takes a bloodmeal from a *T. brucei*-infected mammalian host, the trypomastigote pleomorphic bloodstream forms (BSFs) reach the midgut of the tsetse. While the BSF 'slender' (SL) forms are not capable of surviving in the tsetse, the 'stumpy' (ST) forms are able to differentiate into the trypomastigote procyclic form (PF) (Matthews, 2005; Vickerman, 1985). Within the first week, PFs establish a chronic infection in the tsetse MG and reach the ectoperitrophic space (ES) by a mechanism that remains unclear but assumed to be due to crossing of the tsetse

peritrophic matrix (PM) (Roditi, 1989); see also Chapter 6). During the following week, PFs migrate to the tsetse cardia or proventriculus where the elongated mesocyclic trypomastigotes establish an infection and differentiate into long epimastigote forms (EMFs) (Sharma et al., 2009). These long EMFs undergo an asymmetric division to produce short EMFs (Sharma et al., 2008; Van Den Abbeele et al., 1999). It has been suggested that the long EMFs, still attached to the short ones, escape the PV *via* the oesophagus to migrate towards the salivary glands (SG). Once in the SGs, the long EMFs die whereas the short EMFs attach to the SG epithelium microvilli through the flagellipodia and establish a local SG infection after three weeks post-infection. Attached EMFs can either replicate or eventually differentiate into pre-metacyclic (P-MCF) forms which detach from the epithelium (Rotureau et al., 2012; Tetley et al., 1987; Van Den Abbeele et al., 1999). It has been reported that *T. brucei* may undergo sexual genetic exchange in SG stages only, particularly in rare promastigote-like forms with multiple nuclei and kDNA and expressing meiotic markers (Gibson and Stevens, 1999; Peacock et al., 2014). P-MCFs dwelling free in saliva finally differentiate into non-dividing mature metacyclic trypomastigote forms (MCF) which are the only mammalian-infectious stage in the tsetse (Rotureau et al., 2012). Highly infectious MCF will be injected along with saliva into the mammalian host when the tsetse takes a bloodmeal.

Since tsetse flies are pool feeders (i.e. they break the capillaries to form pools of blood in the skin from where they feed) (Gibson et al., 2017), MCFs are normally injected into the host's dermis and not into the bloodstream directly, from where they first migrate to the lymph nodes and posteriorly to the peripheral bloodstream. Recent studies have shown that MCFs may also reside and proliferate in the dermis at the bite site, and closely interact with adipocytes from the connective tissue (Caljon et al., 2016). This local dermal infection leads to a slight increase in the skin surface temperature as a mechanism to improve transmission. It has been proved to significantly attract more tsetse at the bite site where there are infectious resident parasites.

During the stage 1 of infection, the bloodstream forms populate the blood and lymph of the host. The SL forms replicate to increase parasitaemia and undergo antigenic variation to evade the host immune system (Horn, 2014; Mugnier et al., 2016). When a certain SL density threshold is reached, they are thought to excrete the 'stumpy' induction factor (SIF) to induce the formation ST forms (Mony and

Matthews, 2015; Reuner et al., 1997), although a quorum sensing-independent mechanism for ST formation linked to antigenic variation has been recently described (Zimmermann et al., 2017). STs are known to be arrested quiescent cells that do not undergo antigenic variation and are pre-adapted to be infective to tsetse. STs can be molecularly distinguished from SLs as the former express large amounts of Protein Associated with Differentiation (PAD) transporters, whose surface expression increases as a response to the cold-shock suffered by STs inside the tsetse midgut (Dean et al., 2009). These proteins transport citrate/*cis*-aconitate, which are metabolites that trigger differentiation into PF.

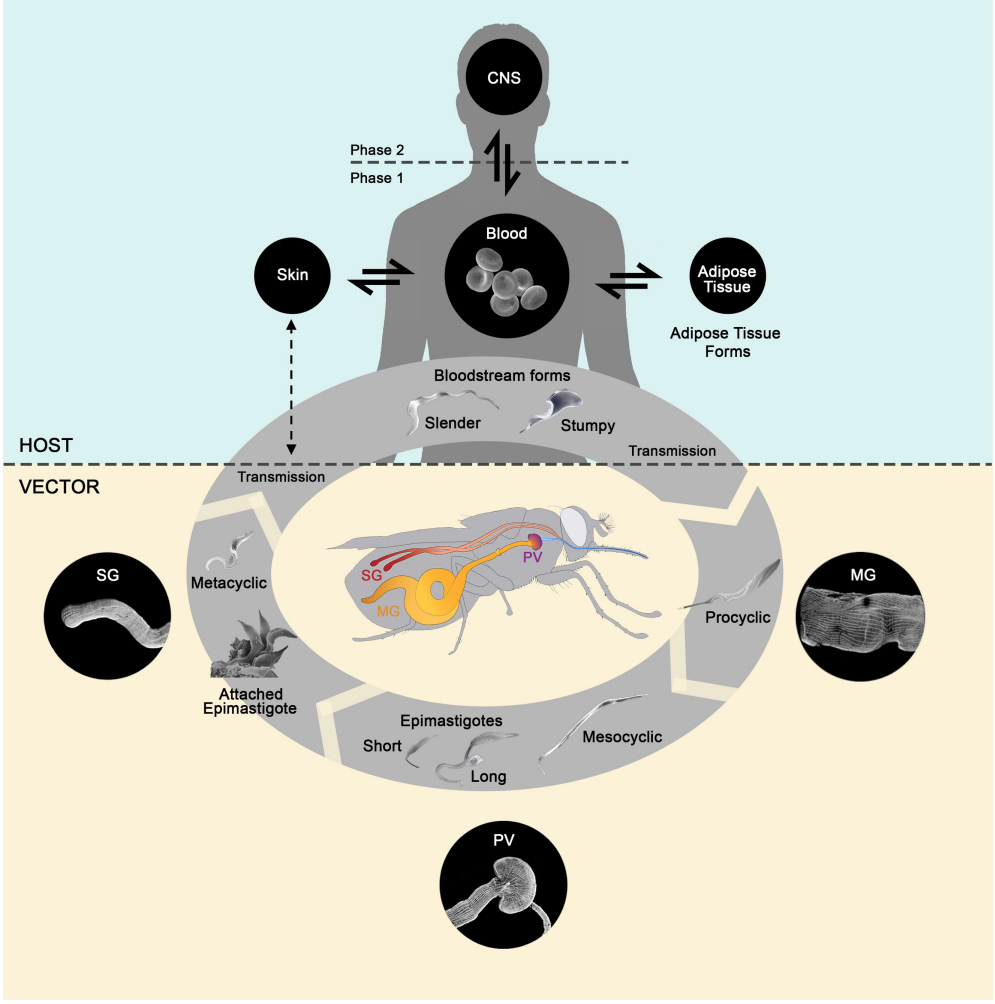


Figure 1.3. *T. brucei* life cycle. *T. brucei* alternates its life cycle between a mammalian host (top) and the tsetse vector. Black circles represent the host spaces *T. brucei* invades which are depicted with actin-stained reconstructed tissues (tsetse) or scanning electron microscopy (host blood). Parasite life stages are represented by

scanning electron microscopy modified images. In the tsetse cartoon, the relevant tissues for trypanosomes infection; i.e. midgut (MG), proventriculus (PV) and salivary glands (SG) are highlighted. The discontinued arrow represents the transmission of metacyclics to the host skin and the possible re-infection of tsetse from skin parasites. The human clinic phases 1 and 2 are represented in the host by the invasion of the central nervous system (CNS).

For many years, it has been understood that BSF remain in the host's bloodstream and lymph, and can eventually undergo extravasation to the liver, spleen and brain (Blum et al., 2008; Kennedy, 2004). It has been recently found that *T. brucei* has, at least, two mechanisms to escape from the hostile environment that are the blood and lymph. The parasites can migrate to the adipose tissue (where they are called adipose tissue forms (ATFs)) and reach high densities comparable to that in blood (Trindade et al., 2016). The ATFs can normally replicate and become STs. However, their overall gene expression is remarkably different from those in blood since they seem to rely on the catabolism of the host's fatty acids through beta-oxidation (Trindade et al., 2016). Alternatively, *T. brucei* can hide into the host skin in where it can replicate, differentiate into ST and be transmitted to the tsetse (Capewell et al., 2016). Importantly, the discovery of these novel human tissue reservoirs challenges the current epidemiological concept of *T. b. gambiense* HAT explained in section 1.2.6 (Casas-Sanchez and Acosta-Serrano, 2016).

1.4 *Glossina* biology

Tsetse flies (*Glossina* spp.) are insects from the order Diptera that populate most territories in sub-Saharan Africa. Their biology and life cycle are particularly unusual compared with most insects. Both male and female feed exclusively on animal blood at least twice a week. The feeding host range is broad including humans, wildlife and livestock. Behavioural studies have shown that they sense and require the host's heat to guide landing (Chappuis et al., 2013). They are attracted to the carbon dioxide exhaled by animals, and some are attracted to blue and black colours (Lindh et al., 2012). After host seeking, the fly lands and penetrates the host skin with the proboscis. The bite breaks the dermis capillaries to form an intradermal pool of blood from where the fly drinks (Gibson et al., 2017). During bloodmeal acquisition,

the fly injects saliva rich in factors that inhibit the haemostatic response to the cutaneous trauma (Alves-Silva et al., 2010). The fly rapidly engorges by temporarily storing the blood into the crop, which then slowly empties into the PV *en route* to the midgut where digestion occurs. Once in the tsetse midgut, the bloodmeal gets compartmentalised by the acellular chitin-rich (Lehane et al., 1996) PM. The tsetse PM (type II) is constantly secreted by epithelial cells in the PV at an estimated rate of ~ 1mm/h in young flies and ~0.44 mm/hour in adults (Lehane and Msangi, 1991), and its main function is to prevent the blood, digestion products and oral pathogens from direct contact with the midgut epithelial cells. The cavity formed between the PM and the gut epithelium is known as ectoperitrophic space (ES), which is assumed to be the preferred place where PF trypanosomes establish a MG infection (Gibson and Bailey, 2003).

After a single mating, the female tsetse stores the sperm in the spermathecae. The sperm are released during each gonadotrophic cycle to fertilise the ovulated egg during the entire life of the female (Langley, 1993). Since tsetse are viviparous, a single larva develops in the uterus after fertilisation, which is nourished by milk produced by dedicated, intrauterine milk glands. The larva is birthed ~9 days after fertilisation and then buries itself into the soil (to avoid desiccation) where it rapidly transforms into a pupa to develop in the dark for 3-4 weeks depending on ambient temperatures. After hatching, the newly emerged fly crawls through the soil back to the surface and within hours can fly and engage in host-seeking (Jackson, 1946). The females reach sexual maturity and deposit the first progeny after 19-22 days under ideal climatic conditions. Ovulation cycles occur every 9-12 days, implying that a female can produce 8-12 larvae during her lifetime.

Tsetse are naturally highly refractory to a trypanosome infection, but age is a critical factor and young, newly emerged flies show the highest susceptibility (Van Hoof, 1937). This is called the teneral phenomenon (Haines, 2013). While adult flies are rarely susceptible to parasite infection, teneral (unfed, recently-emerged flies) tsetse are very susceptible to trypanosome infection mainly due to three factors related with an underdeveloped immunity (Wijers, 1958). First, the PM is yet immature and discontinuous (Wigglesworth, 1929), which facilitates trypanosome invasion of the ES, critical for the establishment of a MG infection. Second, several immune molecules (e.g. tsetse-EP, antimicrobial peptides and ROS), which are known to prevent trypanosome infections, are less active in a teneral MG and PV compared to

adult flies (Haines et al., 2010; Weiss et al., 2013). Lastly, the type of microbiota present at the time of the infection also effects the susceptibility of teneral flies to trypanosome infections (Wang et al., 2013). There are three important tsetse bacterial symbionts which are known to modulate both tsetse and trypanosome physiology. The gram-negative bacterium, *Wigglesworthia glossinidia,* is an obligate symbiont mainly found in the tsetse bacteriome and milk glands, which provides nutritional, immunological and reproductive advantages to the tsetse (Chen et al., 1999). The gram-negative facultative bacterium, *Sodalis glossinidius,* exists intra- and extracellularly in a broad range of tsetse tissues (Cheng and Aksoy, 1999) and has been linked to increase susceptibility to a trypanosome infection by interfering with trypanocidal mechanisms (Dr. Lee Haines, unpublished). Lastly, the gram-negative facultative *Wolbachia* is only located intracellularly in germ line tissues and modulates tsetse fertility. The presence of *Wolbachia* has been found to correlate with an increased resistance to trypanosomes (Cheng et al., 2000).

1.5 *Trypanosoma brucei* major surface glycoproteins

1.5.1 Variant Surface Glycoproteins and antigenic variation

During its life cycle, *T. brucei* expresses abundant surface glycosylphosphatidylinositol (GPI)-anchored glycoproteins, which are developmentally regulated and involved in parasite protection. The best studied of these molecules are the VSGs, which in the blood stages are primarily responsible for antigenic variation (Horn, 2014; Mugnier et al., 2016). Each BSF cell expresses on the surface ~5 x 10^6 VSG homodimers (Cross, 1975), which creates a macromolecular barrier that masks invariant surface proteins such as receptors or transporters. VSG monomers have an estimated molecular size of 50-60 kDa, they are formed by a large N-terminus domain and one or two small C-terminus domains interconnected by a linker domain (**Figure 1.4-A**) (Freymann et al., 1990). Recent studies have shown that these linkers provide great flexibility between domains, so VSGs can adopt two different spatial conformations to adapt the coat thickness in response to the protein density of invariant surface proteins (Bartossek et al., 2017). The GPI anchor of VSG molecules provides free diffusion of these molecules along the membrane (Ferguson et al., 1988). This is essential for the clearance of anti-VSG antibodies by endocytosis of the complex immunoglobulin-VSG and subsequent

recycling of the VSG (Engstler et al., 2007). Most VSGs are known to be *N*-glycosylated with up to three *N*-glycans and, despite *N*-glycans seem not to be essential to create an effective coat, they have been suggested to play a role in filling the empty spaces between homodimers (Mehlert et al., 2002). Besides their role in antigenic variation in the vertebrate host, *T. brucei* VSGs have been recently suggested to promote the colonisation of the tsetse midgut by trypanosomes by downregulating the expression of genes involved in PM synthesis (Aksoy et al., 2016).

T. brucei has a wide repertoire (>1,000) of genes encoding for VSG proteins with variable sequences (Berriman et al., 2005; Cross et al., 2014), but conserved tertiary structures (Blum et al., 1993). Antigenic variation is based on the expression of clone-specific single VSG in combination with periodical switching of the expressed VSG (1 switch/10^5 cells per generation), which is controlled through a mechanism of monoallelic expression that remains poorly understood (Horn, 2014). Surface exposure of VSGs triggers the generation of trypanolytic anti-VSG antibodies by the host's immune system as the major mechanism for parasite clearance (Mugnier et al., 2015). It takes ~1 week to generate the anti-VSG antibody response, time period in which the parasite population expands without control reaching high densities that leads to the formation of ST forms. Then, anti-VSG antibodies effectively clear most of the clonal population except for a minority of cells that switched to express another immunogenically-different VSG. These clones can freely expand within the next antibody-generation period until the new antibody response is launched (**Figure 1.4-B**).

In the blood stages, VSGs are expressed from one of the ~20 sub-telomeric expression sites (BES) (De Lange and Borst, 1982). The basic structure of a BES consists in an RNA polymerase I promoter (Gunzl et al., 2003) placed at ~ 60 kb upstream the *vsg* gene (**Figure 1.4-C**) (Zomerdijk et al., 1990). In between, there is a series of up to 12 expression site-associated genes (ESAGs), which are polycistronically co-transcribed with the *vsg* (Cully et al., 1985). ESAGs are known to be important in immune evasion (i.e. ESAG4 flagellar adenylate cyclase and *T. b. rhodesiense* SRA) or nutrient acquisition (i.e. ESAG6/7 transferrin receptor and ESAG10 folate transporter) (Gadelha et al., 2015). Most ESAGs have paralog genes in chromosomal regions called Genes Related to ESAGs (GRESAGs). A short 70-bp repeats sequence is found upstream the *vsg* gene to promote *vsg* recombination events (Liu et al., 1983). BSF can switch VSG expression by multiple mechanisms

(Horn, 2014). First, they can shut down the expression of the active BES and switch on the expression of a previously-silenced BES (Glover et al., 2016). Second, the *vsg* gene in the active BES can be replaced by homologous recombination through the 70-bp repeats sequence either from a silent BES or from the sub-telomeric arrays of *vsg* genes usually found in minichromosomes (Cross et al., 2014; Williams et al., 1982). Alternatively, *vsg* genes can recombine with each independently from the 70-bp repeats through short conserved sequences in the open reading frame to generate mosaics (Marcello and Barry, 2007).

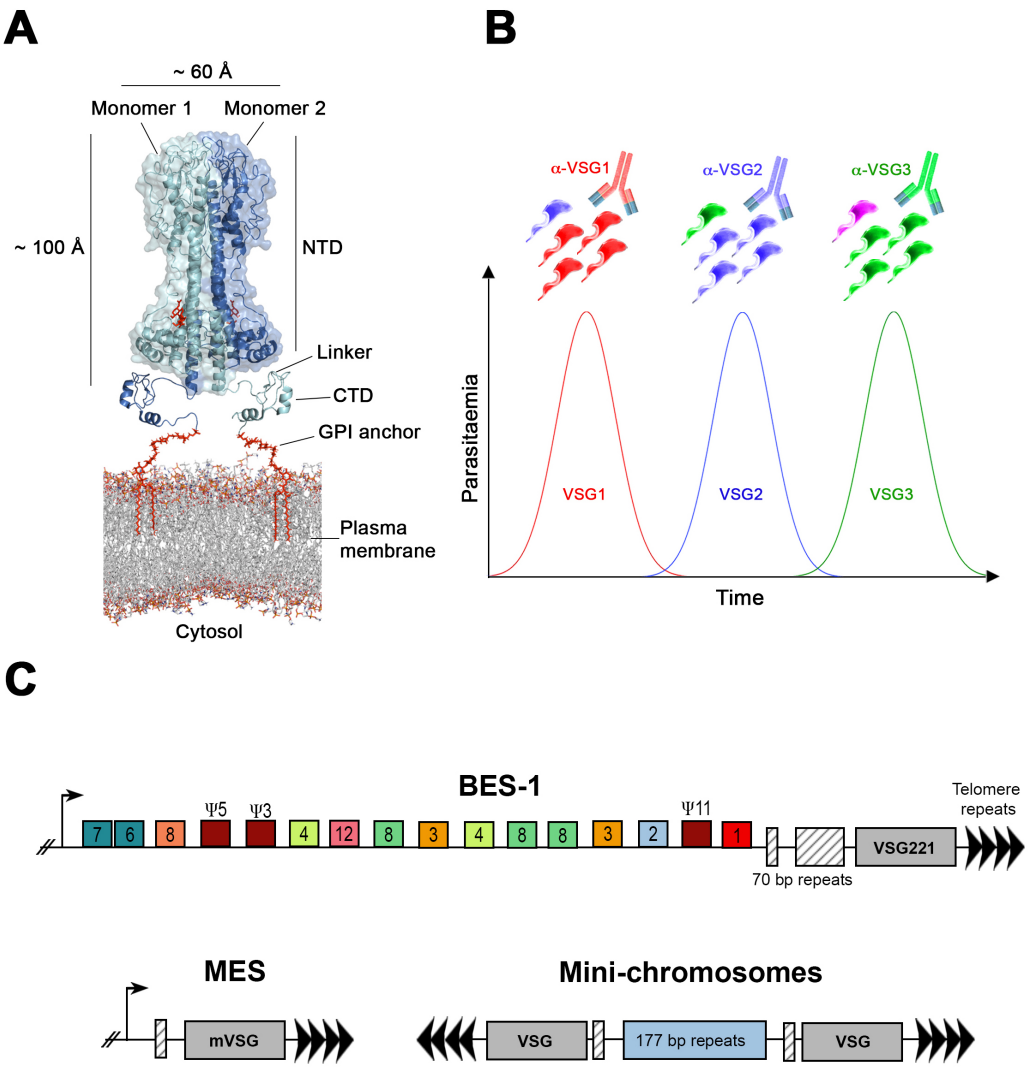


Figure 1.4. VSGs and antigenic variation. (A) Crystal structure of a VSG homodimer (VSG221, PDB: 1VSG) GPI-anchored to the outer face of the BSF plasma membrane. Domains highlighted: N-terminus domain (NTD), C-terminus domain

(CTD), linker and both monomers. **(B)** Schematic of *T. brucei* antigenic variation representing the parasitaemia peaks in function of time in the mammalian bloodstream. **(C)** Structural representation of a BES (BES-1 as example) containing the *vsg221* gene, ESAGs (numbered boxes, Ψ represents the pseudogenes), the 70-bp repeats sequences (striped boxes) and the RNA pol I promoter. Below, a putative MES and a mini-chromosome are represented.

1.5.2 Metacyclic Variant Surface Glycoproteins

The expression of BSF VSGs is shut down right after trypanosomes are taken up by the tsetse vector and start to differentiate into the procyclic stage. Within 24h, the VSG coat gets completely replaced by a new one mainly composed of procyclins (Roditi et al., 1989). However, before the formation of mature MCFs in the SGs, a new coat of metacyclic VSGs (mVSG) is reexpressed by P-MCF parasites (Tetley et al., 1987). Although MCFs do not appear to undergo antigenic variation probably due to their short half-life in the mammalian host, it seems that acquisition of the mVSG coat may still be essential for parasite virulence and survival in the vertebrate host (Tetley et al., 1987). Based on monoclonal antibody recognition, it is estimated that a population of tsetse-derived MCFs expresses as many as 27 different mVSGs (also known as metacyclic variant antigen types (mVATs)) (Barry et al., 1983; Le Ray et al., 1978; Lenardo et al., 1984; Turner et al., 1988), although like in BSFs individual cells express only one mVSG at-a-time. Some studies, in an attempt to develop a vaccine against sleeping sickness, tried to perform immunisations with neutralised MCF to generate protection against all mVATs. Despite obtaining some level of protection against most mVATs, full protection was never achieved. mVSGs are expressed from the metacyclic expression sites (MES) placed in sub-telomeres of the largest chromosomes (Alarcon et al., 1994; Barry et al., 1983) (**Figure 1.4-C**). Unlike BES, MES lack functional ESAGs. There is no apparent recombination between mVSGs because they lack the 70-bp repeats sequences upstream the *mvsg* gene. Remarkably, Pol-I transcription of MES is monocistronic and life-stage regulated rather than post-transcriptionally as the vast majority of *T. brucei* genes (Graham and Barry, 1995; Lenardo et al., 1986) (see section 1.6). In general, MCFs have been under-investigated compared to BSFs due to the scarce yield of cells obtained from inefficient controlled tsetse infections, and to the lack of an *in vitro* production system. Thus, it has been assumed over decades that the surface glycocalyces of both MCFs

and BSFs are equivalent and that in both, their VSG molecules hide the low copy number invariant surface proteins from the immune system.

1.5.3 Procyclins

PFs are known to abundantly express GPI-anchored procyclins on the surface, a family of proteins that appears important for an efficient cyclical transmission in the tsetse, but whose biological function remains unknown (Vassella et al., 2009). Within the first days of MG infection, recently-differentiated or 'early' procyclic cells first express GPEET, a form of procyclins that contains 5-6 Glu-Pro-Glu-Glu-Thr repeats at the C-terminus and in which all Thr residues can be phosphorylated (Mowatt and Clayton, 1987; Roditi and Clayton, 1999; Roditi et al., 1989). After several days, GPEET gets then replaced by a family of EP procyclin isoforms, which contain 18-30 Glu-Pro repeats at C-terminus depending on the isoform, and whose expression is maintained in trypanosomes infecting either the tsetse MG or PV (Roditi et al., 1987; Sharma et al., 2008). There are several EP isoforms (EP-1, EP-2 and EP-3) that differ in the length of the repeats domain and the presence or absence of *N*-glycans (Acosta-Serrano et al., 1999; Acosta-Serrano et al., 2001)(see section 1.7.5). In culture, GPEET expression can be maintained in the presence of glycerol and absence of glucose, otherwise PCF preferentially express EP isoforms (Vassella et al., 2000). Procyclins transcription is carried out by the RNA polymerase I in the nucleolus (Schumann Burkard et al., 2013). Remarkably, both GPEET and EP procyclins undergo N-terminal proteolysis in the tsetse MG but not in culture, suggesting that cleavage is probably produced by vector proteases (Acosta-Serrano et al., 2001).

1.5.4 Brucei Alanine-Rich Proteins (BARPs)

Further in development, mesocyclic trypomastigotes in the PV still conserve the EP procyclin coat and, although controversial, the asymmetrically dividing PV EMFs still display EP remains (Sharma et al., 2008). Parasites dwelling in SGs, however, despite transcribing procyclin genes, no longer translate them into proteins (Acosta-Serrano et al., 2001; Urwyler et al., 2005). The procyclin surface coat of attached epimastigotes in the SGs is replaced by a coat of BARPs (Urwyler et al.,

2007). There are 14 BARP isoforms in the *T. brucei* genome (Berriman et al., 2005), and their expression is up to 20-fold upregulated in SG parasites compared to MG procyclics (Urwyler et al., 2007). The structure and biological role(s) of BARP remain unknown to date. *T. brucei* BARP haven been suggested to be analogue proteins to the *T. congolense* Glutamic Acid-Rich proteins (GARPs) (Bayne et al., 1993), which are also expressed in epimastigote forms infecting the tsetse mouthparts, and their function remains undetermined.

1.6 Gene expression regulation in *Trypanosoma brucei*

In *T. brucei*, long arrays of genes are transcribed polycistronically by the RNA polymerase II (Berriman et al., 2005; Johnson et al., 1987), except for major surface glycoproteins such as VSGs or procyclins, which are transcribed by RNA polymerase I (Gunzl et al., 2003; Schumann Burkard et al., 2013). The polycistronic transcripts become then monocistronic after a *trans*-splicing event coupled to polyadenylation of the 3' ends (**Figure 1.5**) (LeBowitz et al., 1993; Matthews et al., 1994). During *trans*-splicing, a 39-nucleotide splice leader transcribed by the RNA pol II or III, containing the 5' cap structure required for translation, is added at the 5' end of each spliced monocistronic transcript, and individual transcripts are then separated by cleavage from upstream genes by RNAses (Patzelt et al., 1989; Perry et al., 1987). *cis*-splicing is virtually absent in *T. brucei*. Unlike higher organisms, gene expression in *T. brucei* is mainly regulated post-transcriptionally at the level of *trans*-splicing and polyadenylation, which determine RNA stability (Liang et al., 2003). However, mature mRNA levels may greatly vary between flanking genes or within the same gene during different developmental stages of the parasite (Archer et al., 2009; Vasquez et al., 2014). This precise regulation, which shapes most of the morphological and metabolic changes, essential for parasite adaption to the different environments, is mainly carried out by RNA-binding proteins (RBPs) (Clayton, 2013; Glisovic et al., 2008). RBPs have been shown to be involved in life cycle development, differentiation, cell cycle, rRNA processing and heat shock response. The *T. brucei* genome contains ~150 genes encoding for proteins with putative RNA-binding domains (Berriman et al., 2005). RBPs typically bind the untranslated regions (UTRs) of the transcripts to modulate their fate, including synthesis, localisation, turnover and decay among others (Clayton, 2013; Siegel et al., 2011). For instance, RBP10 is known to play a role in BSF differentiation as overexpression in PCF led to the reverse differentiation

to BSF (Mugo and Clayton, 2017; Mugo et al., 2017). On the other hand, RBP6 which is only expressed in PV trypanosomes, was found to be related with differentiation too when its overexpression forced PCF to differentiate into EMFs and MCFs *in vitro* (Kolev et al., 2012).

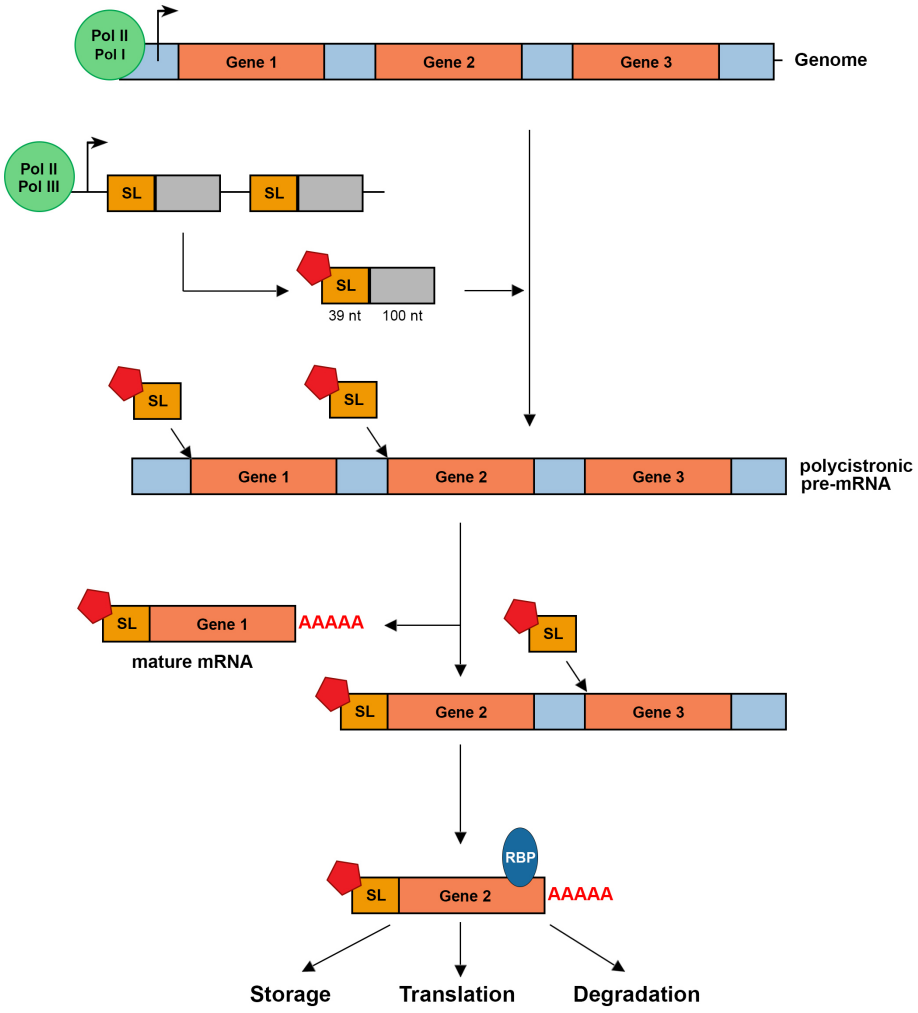


Figure 1.5. Post-transcriptional regulation of gene expression in *T. brucei*. An array of chromosomal genes is transcribed by the RNA pol I or II to create a polycistronic pre-mRNA. Simultaneously, the splice leader (SL) 39 nt sequence is transcribed by the RNA pol III or II and modified by the addition of a 5' cap and depletion of a 100 nt 3' sequence. *Trans*-splicing of the polycistronic pre-mRNA couples the SL to single monocistronic mRNAs which are polyadenylated. The RNA-binding proteins (RBPs) can then bind the 3' UTR of the mRNA and determine its fate involving either translation to protein, storage or degradation.

1.7 *N*-glycosylation

Virtually, all eukaryote organisms are able to undergo protein glycosylation, even ancient organisms like *Plasmodium* spp. which have truncated glycosylation pathways (Cova et al., 2015). Protein glycosylation is a post-translation modification (PTM) consisting in the attachment of mono-, oligo- or polysaccharide to a newly synthesized protein (Parodi, 1998; Varki, 1993). To date, 5 different types of protein glycosylation have been described in eukaryote organisms, among which *N*-glycosylation and *O*-glycosylation are the most common ones. *N*-glycosylation consists in the addition of an *N*-glycan chain to the sequon Asn-X-Ser/Thr (where X can be any residue except proline) (Marshall, 1972). Among other properties, *N*-glycans are important for protein folding, stability and solubility (Helenius and Aebi, 2004). However, the role of *N*-glycans may vary between proteins and glycosylation sites, and their depletion may not always critically affect the normal function of the protein (Kornfeld and Kornfeld, 1985). *N*-glycans are attached to proteins *via* the anomeric carbon of the innermost β-*N*-acetylglucosamine (GlcNAc) residue linked to the Asn amide nitrogen (Marshall, 1972). The sole presence of the Asn-X-Ser/Thr sequon does not always ensure the addition of the *N*-glycan as the residues surrounding the sequon have been shown to play a role in the processing (Izquierdo et al., 2009b). Addition of *N*-glycans to proteins is the main type of protein glycosylation present in African trypanosomes, including all *T. brucei* sub-species (Ferguson, 1997) .

1.7.1 *N*-glycan structures

There are three main types of *N*-glycans based on their structure and composition: oligomannose, complex and hybrid *N*-glycans (**Figure 1.6**) (Varki et al., 2015). They all share a common core structure of Manα1-6(Manα1-3)Manβ1-4GlcNAcβ1-4GlcNAcβ, but differ in the extensions. The oligomannose type *N*-glycans typically contain two to six α-mannoses on top of the three mannose residues from the core (five to nine mannose residues in total) thus creating a tri-antennary glycan. In contrast, the complex type *N*-glycans have a more diverse composition typically including bisections from the core mannose residues of NeuAcα2-3/6Galβ1-4GlcNAcβ (where Neu is sialic acid and Gal is galactose). Other common variants

include tri- and tetra-antennary structures. Hybrid *N*-glycans present characteristic structures from both oligomannose and complex types. They usually contain mannose residues attached to the Manα1-6 arm and a bi- or tri-antennae containing a complex type structure attached to the Manα1-3 arm (**Figure 1.6**).

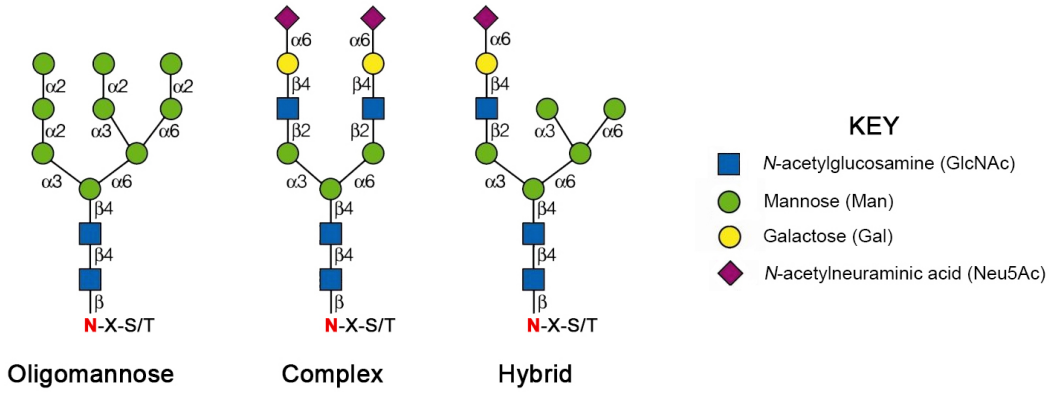


Figure 1.6. *N*-glycan structural types. Schematic of the representative structures for the three major types of *N*-glycans: oligomannose, complex and hybrid. Monosaccharide code on the right, linkages annotated between residues.

1.7.2 *N*-glycan biosynthesis and processing

The biosynthesis and processing of *N*-glycans occur entirely in the ER and Golgi apparatus of the eukaryote cells, carried out by a series of enzymes associated to the membranes of these organelles (**Figure 1.7**) (Helenius and Aebi, 2004; Helenius et al., 2002; Kornfeld and Kornfeld, 1985). Briefly, a precursor lipid-linked oligosaccharide is first synthesised and posteriorly transferred to the glycoprotein in the ER. The resulting (immature) *N*-glycan is usually trimmed and modified in both the endoplasmic reticulum (ER) and Golgi apparatus to produce its final structure.

The formation of the lipid-linked precursor oligosaccharide starts in the cytoplasmic face of the ER membrane. Two distinct transferases, UPD-GlcNAc:dolichol-*P* (ALG7) and UDP-GlcNAc:GlcNAc-1-*P* (ALG13/14), add two GlcNAc residues to dolichol-phosphate (*P*-Dol) –from the nucleotide activated sugar donor uridine diphosphate (UDP)-GlcNAc– forming $GlcNAc_2$-*P*-*P*-Dol. Posteriorly, a

series of mannosyltransferases (ALG1, ALG2 and ALG11) add 5 mannose residues –using Man-*P*-Dol as donor– forming $Man_5GlcNAc_2$-*P*-*P*-Dol. This precursor, still facing the cytosolic side of the ER membrane is then translocated into the luminal ER face by a flippase known as RFT1 (Helenius et al., 2002) thus becoming the substrate of the mannosyltransferases ALG3, ALG9 and ALG12 forming $Man_9GlcNAc_2$-*P*-*P*-Dol. Unlike Man-transferases involved in the synthesis of $Man_5GlcNAc_2$-*P*-*P*-Dol, ALG3, ALG9 and ALG12 use GDP-Man as donor. Depending on the organism and cell type, the intermediate $Man_9GlcNAc_2$-*P*-*P*-Dol is further decorated with three α-glucose (Glc) residues forming the final mature precursor $Glc_3Man_9GlcNAc_2$-*P*-*P*-Dol. The glucosyltransferases, ALG6, ALG8 and ALG10, which use GDP-Glc as donor, are in responsible for transferring these three terminal Glc residues.

The mature oligosaccharide is co-translationally transferred *en bloc* to the nascent glycoprotein by the oligosaccharyltransferase complex (OST). The mammalian OST complex is formed by up to nine transmembrane proteins, being STT3 the catalytic subunit. After transfer, the *N*-glycan is trimmed in the ER lumen to remove the three Glc residues and at least one Man residue by the exoglycosydase glucosidase I/II and the ER mannosidase I, respectively. The *N*-glycan is then monoglucosylated by a glucosyltransferase to enter into the calnexin/calreticulin (chaperones) cycle for proper protein folding and quality control (Helenius and Aebi, 2004; Parodi, 2000). Incorrectly folded proteins can either lose the glycan, be re-glucosylated and re-folded, or signalled for degradation through a process known as endoplasmic-reticulum-associated protein degradation (ERAD). Unlike lysosomal proteins, most glycoproteins leave the ER and enter the Golgi apparatus for further glycan processing. At the Golgi, mannosidase I removes up to four Man residues and the resulting glycan (oligomannose) can then remain unaltered or can be further modified by a broad range of enzymes to form complex and hybrid type *N*-glycans. The final structure of a particular *N*-glycan is not directly encoded in the DNA code of a gene but depends on multiple factors related with glycan processing in the Golgi, including availability and concentration of activated sugar donors, abundance and compartmentalisation of transferases, glycan accessibility, and differential speed of the reactions, among others.

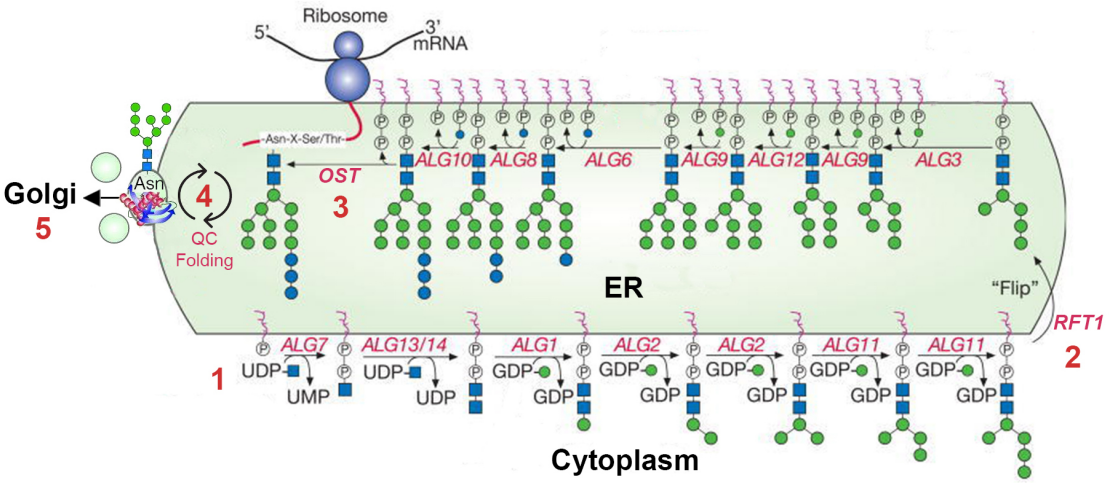


Figure 1.7. *N*-glycosylation synthesis pathway. Assembling of the precursor oligosaccharide starts (1) at the ER cytosolic face where a series of enzymes (pink italics font) build up the $Man_5GlcNAc_2$-*P*-*P*-Dol which is then translocated into the luminal ER face (2) where the structure is modified until the final precursor $Glc_3Man_9GlcNAc_2$-*P*-*P*-Dol is made. The precursor is then linked *en bloc* to an asparagine from the nascent glycoprotein by the oligosaccharyltransferase (OST). The *N*-glycoprotein enters the calnexin/calreticulin cycle for quality control of folding (4), the *N*-glycan structure is reduced to a $Man_9GlcNAc_2$ and the folded *N*-glycoprotein exported to the Golgi apparatus (5) for further trimming and modifications of the *N*-glycan. Schematic modified from Essentials of Glycobiology, A. Varki.

1.7.3 *N*-glycosylation in *Trypanosoma brucei*

T. brucei has a simpler *N*-glycosylation machinery compared to most of eukaryotes, but efficient processing (~96%) of all potential *N*-glycosylation sites (Parodi, 1993a, b). Trypanosomes cannot synthesize the activated sugar donor, Glc-*P*-Dol, and also lack the glucosyltransferases ALG6, ALG8 and ALG10. Consequently, they are unable to undergo the glucosylation steps to form the mature precursor $Glc_3Man_9GlcNAc_2$-*P*-*P*-Dol from $Man_9GlcNAc_2$-*P*-*P*-Dol. The lack of glucosylation of the *N*-glycan precursor is characteristic of all trypanosomatid parasites. *T. brucei* in particular can either transfer $Man_9GlcNAc_2$ or $Man_5GlcNAc_2$, depending on the parasite life stage; i.e. while BSFs transfer both glycans, procyclics only transfer $Man_9GlcNAc_2$. The quality control calnexin/calreticulin is present in trypanosomatids (despite having unglucosylated precursor oligosaccharides) thanks

to a glucosyltransferase dependent on uracil diphosphate(UDP)-Glc and a glucosidase II. In addition, the OST complex is much simpler compared to that from mammals and it is only composed by the catalytic subunit STT3.

1.7.4 Glycosylphosphatidylinositol (GPI) anchors

GPI anchors are PTMs that allow the attachment of proteins to the outer leaflet of the plasma membrane through a lipid-linked glycan, as an alternative to protein transmembrane domains (Ferguson, 1999). GPI anchors are ubiquitous among eukaryotes and especially abundant in protozoa such as kinetoplastids (Ferguson and Williams, 1988), being the only exception so far *Trichomonas vaginalis*, which lacks the whole GPI machinery (Hirt et al., 2011). The first GPI anchor structure was solved from a *T. brucei* VSG (Ferguson et al., 1988). GPI-anchored proteins may have very diverse functions such as adhesion, signalling, hydrolysis or cell protection. All known GPI anchors are attached through an ethanolamine phosphate (EtNP) group to the C-terminus of the protein, and they are formed by a common core structure (Varki et al., 2015). The oligosaccharide Manα1-2Manα1-6Manα1-4GlcNα1-6 (being GlcN glucosamine) links the EtNP to the phosphatidylinositol (PI) moiety containing a D-*myo*-inositol-1-P and up to three fatty acids that anchor the molecule to the plasma membrane (McConville and Ferguson, 1993). GPI anchor structures are broadly variable since up to eight side chains of the oligosaccharide can be substituted by several chemical groups, and the glycerol-fatty acids may vary in length and chemistry, including the loss of one fatty acid forming a *lyso*-PI. GPI anchors are first synthesized in the ER, in where they are posteriorly attached to the newly-synthesized protein with simultaneous cleavage of the GPI anchor signal peptide. Lastly, both the fatty acids and the carbohydrate side chains can be modified in the ER and Golgi.
Some protozoa and mycobacteria also express glycoinositol phospholipids (GIPLs) on the cell surface resembling protein-free GPI anchors (Ferguson, 1997). They have been reported in *T. brucei*, *T. cruzi* and *L. major* among others, and their biological function remains undetermined.

1.7.5 *N*-glycan and GPI anchor structures in *Trypanosoma brucei*

In *T. brucei*, *N*-glycosylation has been found to be the only type of protein glycosylation. *N*-glycan and GPI anchor structures have been extensively studied in major surface glycoproteins of biological relevance, mainly VSGs and procyclins (**Figure 1.8**) (Ferguson, 1997). In addition, in *T. brucei* BSFs, flagellar pocket glycoproteins can be modified with giant complex-type *N*-glycans that can contain more than 50 *N*-acetyl-lactosamine (Galβ1-3GlcNAc) repeats (Atrih et al., 2005). These structures are the largest *N*-glycans ever reported from any organism.

GPEET procyclins lack *N*-glycosylation sites, and only EP-1 and EP-3 have been found to be *N*-glycosylated in a single N-terminal residue (Asn^{29}) (Acosta-Serrano et al., 1999; Hwa et al., 1999). The only *N*-glycan identified in these sites is the oligomannose $Man_5GlcNAc_2$. Procyclin GPI anchors (either from GPEET or EP) show more structural variability and complexity. The PI moiety is unusual in that it is composed of *lyso*-(acyl)PI species, containing mostly stearic acid (C18:0) in the *sn*-1 position and a heterogeneous acyl chain (mainly C18:0, C18:1 and C18:2) linked to the inositol ring. Lastly, the core mannoses are decorated with branched polylactosamine repeats that may be sialylated, forming oligosaccharides of average composition $NeuAc_5Gal_9GlcNAc_9$ (Izquierdo et al., 2009a; Mehlert et al., 1998). Since *T. brucei* cannot synthesise sialic acid, this one is added by trans-sialidases depending on the presence of the sugar donor in the environment (from either the mammalian host or tsetse vector) (Engstler et al., 1993; Engstler and Schauer, 1993).

VSGs, in contrast, can present more complex *N*-glycans but simpler GPI anchor structures. The main types of *N*-glycan found linked to VSG polypeptides are the oligomannose $Man_9GlcNAc_2$ and variable complex-type glycans that may contain terminal αGal residues (i.e. Galα1-3)$_2$(Galβ1-4GlcNAc)$_2$$Man_3GlcNAc_2$ (Ferguson, 1997) and or several degrees of poly-acetyllactosamine repeats (Galβ1-4GlcNAc)$_n$$Man_3GlcNAc_2$. As for the composition of the VSG GPI anchor, it contains a lipid moiety exclusively composed of di-myristoyl(C14:0)glycerol-PI, while the mannosylated glycan core is decorated with a variable number of αGal residues, depending on the VSG type (Ferguson, 1997; Izquierdo et al., 2009a; Mehlert et al., 1998).

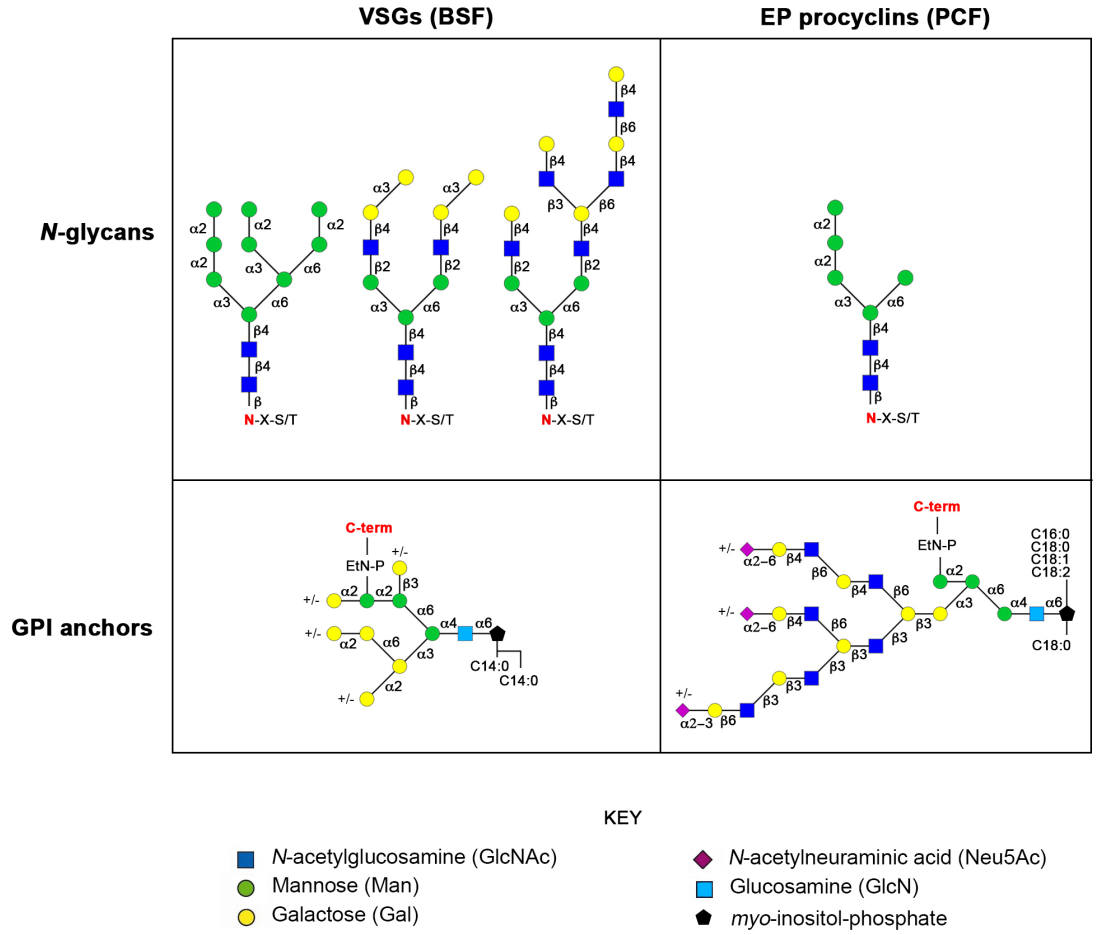


Figure 1.8. *N*-glycan and GPI glycan structures in *T. brucei*. Major structures identified in VSGs from BSF (left) and EP procyclins from PCF. Monosaccharide key at the bottom, linkages annotated between residues. The +/- symbols indicate the residues that can be sometimes absent. Sialic acid is added depending on the expression of trans-sialidases and the presence of the sugar donor in the environment.

The *N*-glycan or GPI anchor structures from other major *T. brucei* surface glycoproteins, such as BARP or mVSGs (from SG EMFs and MCFs, respectively), remain unknown mainly because these life stages cannot be cultured *in vitro*, and tsetse-derived parasites cannot be obtained in sufficient number for standard structural analyses based on mass spectrometry approaches.

1.7.6 Functional studies on *Trypanosoma brucei N*-glycosylation

A key enzyme in the *N*-glycosylation pathway, STT3, has been extensively used for the study of *N*-glycosylation (Castro et al., 2006; Hese et al., 2009; Izquierdo et al., 2009b; Nasab et al., 2008). The *T. brucei* genome contains three paralogs (stt3a, stt3b and stt3c) and one pseudogene (*stt3h*) encoding for at least three distinct STT3 catalytic subunits of the trypanosome OST, which have no homology with OST subunits found in higher organisms (Berriman et al., 2005). It has been determined that *T. brucei* differentially expresses the STT3 isoforms depending on the stage of development (Izquierdo et al., 2009b). Cultured BSF express STT3A and STT3B in a similar proportion, while PCF mainly express STT3B. The expression of STT3C was found to be undetectable in both parasite stages and therefore it has been suggested that it may play a role in the tsetse stages of the parasite (e.g. epimastigotes). Unlike STT3A, both STT3B and STT3C were found to rescue *Saccharomyces cerevisiae* Δstt3 mutants. Interestingly, isoforms STT3A and STT3B have very different specificities regarding the type of oligosaccharide precursor they better transfer. For example, STT3A has a clear preference to transfer the intermediate biantennary precursor $Man_5GlcNAc_2$, which has been proved to be the only route for the formation of complex *N*-glycans in BSFs. In contrast, STT3B only transfers the full triantennary precursor $Man_9GlcNAc_2$ as the sole route for the creation of oligomannose *N*-glycans. Consequently, BSF can *N*-glycosylate proteins with complex and oligomannose glycans, as previously described in VSGs (Mehlert et al., 1998), because they express both STT3A and STT3B, while PCF mainly express STT3B and thus only oligomannose *N*-glycans are found in the procyclins (Hwa et al., 1999).

In addition, the STT3 isoforms also differ in the specificity they present for the amino acid residues flanking the *N*-glycosylation sequon (Izquierdo et al., 2009b). While STT3A and STT3C were found to present similar patterns of specificity for acidic residues, STT3B has an increase tolerance for different flanking residues. Accordingly, the isoelectric point of the flanking residues of the glycosylation sequon can directly determine which isoform will preferentially glycosylate a specific site and with which oligosaccharide precursor, thus determining the *N*-glycan type. However, other factors like secondary and tertiary structure of the polypeptide, or the position that the *N*-glycosylation site occupies in the protein may slightly alter this dogma. As for the essentiality of STT3 in *T. brucei*, partial depletion of *N*-glycosylation after RNAi of *stt3* genes led to severe growth defects in cultured BSF, and this effect was

magnified *in vivo* resulting in low parasitaemia levels in infected mice (Izquierdo et al., 2009b). The essentiality of *N*-glycosylation for development and transmission in the tsetse remains unknown, although *N*-glycans do not seem to be essential for the *in vitro* growth of PCFs. Interestingly, BSF cells exposed to glycan-binding agents which normally induce cell lysis became resistant *in vitro* by either down-regulation of specific *stt3* genes (Castillo-Acosta et al., 2016) or by undergoing recombination between *stt3* genes to create chimera enzymes which produced glycan structures for which the glycan-binding agents had no affinity (Castillo-Acosta et al., 2013). However, mouse infectivity of these mutants was severely compromised, highlighting the importance of specific glycan structures for parasite virulence.

1.8 Aims of the thesis

As reviewed above, vaccine development against sleeping sickness and African animal trypanosomiases has been hampered mainly because of antigenic variation and the masking of invariant surface proteins by the VSG coat in the parasite blood stages. In addition, the inability to culture *in vitro* metacyclic trypomastigotes, together with the difficulties in accessing a suitable tsetse facility for infection experiments, have prevented the development of a transmission-blocking vaccine against trypanosomiasis. Thus, this thesis aimed to identify and characterise novel targets in tsetse-derived metacyclic trypomastigotes for the potential development of transmission-blocking or preventive approaches against African trypanosomiases. The specific aims of this thesis were:

1. To characterise a novel family of *T. brucei* invariant surface proteins herein named as Metacyclic Invariant Surface Proteins (MISP)
2. To establish and characterise an *in vitro* production system of *T. brucei* metacyclic forms for the study of MISP.
3. To identify and characterise *T. brucei* genes whose expression is essential for its development into the metacyclic transmissible stage.
4. To characterise surface *N*-glycans from *T. brucei* metacyclics and the essentiality of *N*-glycosylation for parasite transmission.
5. To develop new methods for imaging *T. brucei* in the tsetse with the aim to study parasite development within the vector.

CHAPTER 2

Materials and Methods

2.1 Cell culture methods

2.1.1 *Trypanosoma brucei brucei* bloodstream forms

T. b. brucei strain AnTat 1.1 90:13 (MacGregor et al., 2013) bloodstream forms were cultured in HMI-9 medium (Hirumi and Hirumi, 1989) (Gibco) supplemented with 10% heat-inactivated faetal bovine serum (FBS), containing G418 (2.5 μg/mL, Invivogen) and Hygromycin B (4 μg/mL, Invivogen) in 25 cm^2 or 75 cm^2 plastic vented non-treated flasks (Corning) at 37°C and 5% CO_2. Cells were typically maintained at exponential growth phase <10^6 cells/mL and passaged every 1-2 days by dilution with fresh medium, to prevent the formation of 'stumpy' non-proliferative forms.

2.1.2 *Trypanosoma brucei brucei* procyclic cultured forms

T. b. brucei strain AnTat 1.1. 90:13 (MacGregor et al., 2013), strain Lister 427 29:13 (Wirtz et al., 1999) and strain TSW-196 (Paindavoine et al., 1986) were cultured in SDM-79 medium (Gibco) supplemented with 10% heat-inactivated FBS, haemin 7.5 mg/L, L-glutamine 320 mg/L and GlutaMax (Gibco) in 25 cm^2 or 75 cm^2 plastic vented non-treated flasks (Corning) at 27°C and 5% CO_2. The parental marker selection drugs G418 (15 μg/mL) and Hygromycin B (50 μg/mL) were permanently added to the medium except for the strain TSW-196. Cells were typically maintained at exponential growth phase (i.e. between 10^6 cells/mL and $2x10^7$ cells/mL) and passaged every 2-3 days by dilution in fresh medium. Alternatively, the strain AnTat 1.1 90:13 was cultured in DTM medium (Ziegelbauer et al., 1990) supplemented with 10% FBS, haemin 7.5 mg/L, L-glutamine 1,6 g/L and GlutaMax (Gibco).

2.1.3 *Trypanosoma brucei brucei* stabilates

T. b. brucei cultures were preserved and stored in a cryostore with liquid nitrogen (-195.7°C). BSF cells (~10^6) or PCF cells (~10^7) were spun down at 1,000 x *g* for 5 minutes at room temperature (RT) to remove the culture media and resuspended in 0.5 mL of HMI-9 10% FBS or SDM-79 10% FBS, respectively, supplemented with 20% glycerol (as cryoprotectant) in a cryovial. Cells were first frozen at -80°C using a Mr. Frosty freezing container (Nalgene, 1°C/min cooling)

before being transferred to the cryostore. To set a new culture from frozen stabilates, cryovials were warmed up at ~37°C, and spun down at 1,000 x *g* for 5 minutes at RT to remove the cryoprotective media. Cells were then resuspended in pre-warmed HMI-9 10% FBS (37°C) or SMD-79 10% FBS (27°C) media according to the cell type and let recover in a flask overnight before passaging.

2.1.4 Induction of the tetracycline-inducible genetic system in *T. b. brucei*

The *T. b. brucei* strains AnTat 1.1 90:13 and Lister 427 29:13 are mutant clonal lines created by the community for the expression of transgenes under the control of a tetracycline (Tet-ON) inducible system (Wirtz et al., 1999). Both strains contain the insertion of the construct pLEW13 (*neo* marker) in the α/β tubulin array for the constitutive expression of the T7 RNA polymerase and the tetracycline repressor (Tet R) (**Figure 2.1**). In addition, they either have the insertion of the pLEW90 (AnTat, in the tubulin array, *hyg* marker) or pLEW29 (Lister 427, in the RNA polymerase I locus, *hyg* marker) for the constitutive expression of a second Tet R gene under the control of a modified (10% of wildtype activity) T7 promoter.

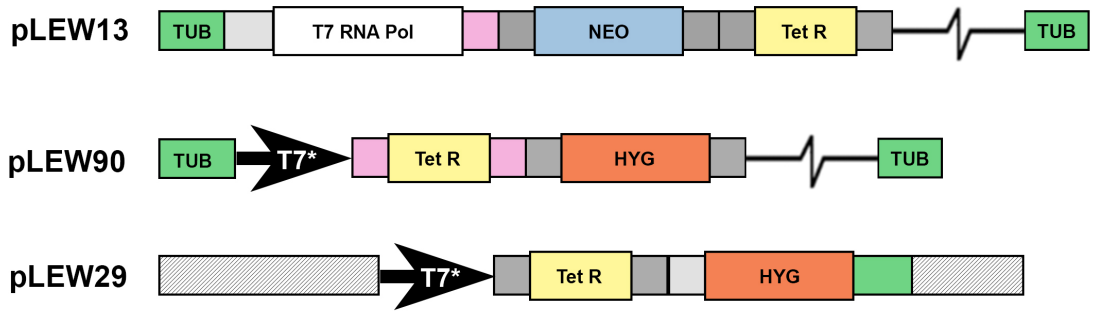


Figure 2.1. Transgene insertions required for the tetracycline-inducible system in the *T. b. brucei* parental lines AnTat 1.1. 90:13 and Lister 427 29:13. The pLEW13 insertion in the tubulin locus provides the T7 RNA polymerase (T7 RNA Pol), the tetracycline repressor (Tet R) and it is selected with the neomycin (NEO) marker gene. The pLEW90 insertion (tubulin locus) or pLEW29 insertion (RNA Pol I locus, striped box) provide and extra copy of Tet R driven by a modified T7 promoter (T7*) and they are selected with the hygromycin (HYG) marker gene. Coding sequences are flanked by 5' and 3' UTRs either from actin (dark grey), aldolase (pink), tubulin

(green) or procyclin (light grey) to provide stability to the transcripts. Adapted from Wirtz et al. 1998.

These genetic backgrounds allow the transcription of genes under the strong viral promoter T7, and most importantly, the tight regulation of gene transcription by the addition or removal of tetracycline (Tet) or its analogue doxycycline (Dox). The gene under tetracycline induction needs to contain the tetracycline operator (Tet Op) DNA sequences (typically 2 to 3 in tandem) between the promoter and the 5' regulatory UTR sequence. By default, the constitutively-expressed Tet R protein strongly binds to the Tet Op sequences and physically blocks transcription of the transgene (**Figure 2.2**). When Tet or Dox are added to the culture, the drugs enter the cell by diffusion and detach the Tet R from the Tet Op sequences because of the higher affinity between them, unblocking the path for the RNA polymerase to transcribe the transgene. The estimated half-life of Tet (2 days) and of Dox (4 days) is relatively short and therefore they need to be continuously added to the culture medium. Otherwise, the drugs can be removed from culture to eventually repress the transgene expression again.

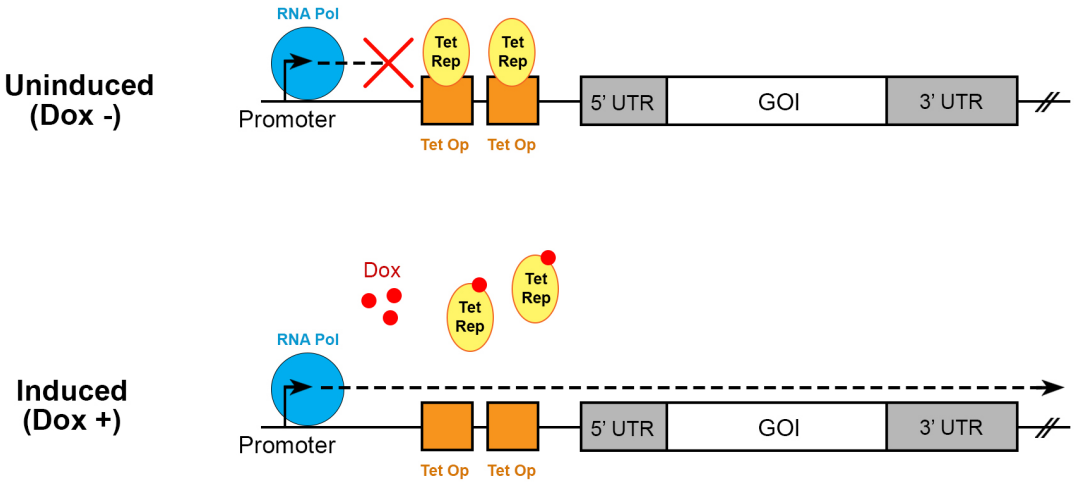


Figure 2.2. Schematic of the tetracycline-inducible system in *T. brucei*. When uninduced (Dox -), the RNA polymerase (RNA Pol, blue circle) binds to the promoter but is unable to transcribe the gene of interest because the tetracycline repressors (Tet Rep, yellow ovals) are bound to the tetracycline operator (Tet Op, orange squares) DNA sequences downstream the promoter. Only when doxycycline (Dox,

red circles) is added the system is induced (Dox +) as Dox binds to the Tet Rep detaching it from the Tet Op DNA sequence and thus allowing the RNA Pol to transcribe the GOI.

For Tet-inducible transgene induction, both BSF and PCF cultures were supplemented with 1 μg/mL doxycycline (1 mg/mL stock in 70% ethanol). Uninduced controls were supplemented with equal volume of 70% ethanol only. Transgene induction can also be applied to trypanosomes infecting the tsetse by supplementing every bloodmeal with 2 μg/mL doxycycline or the same volume of 70% ethanol only as uninduced control. Doxycycline may increase tsetse mortality after 2 weeks of treatment as it may affect the bacterial symbionts which provide essential metabolites to the tsetse (unpublished observation).

2.1.5 *Escherichia coli* culture

E. coli strain DH5α (ThermoScientific), strain Rosetta-gami 2 (DE3) pLysS (Merk Millipore) or strain Rosetta 2 (DE3) pLysS (Merk Millipore) were grown for multiple purposes in autoclaved Lennox lysogenic broth (LB) medium 20 g/L supplemented with selection antibiotics (see **Table 2.1** for antibiotic concentrations) at 37°C and 220 rpm shaking overnight. Alternatively, *E.* coli were grown on LB Lennox (20 g/L) agar (15 g/L) plates supplemented with antibiotics (**Table 2.1**) at 37°C overnight. Plates were kept at 4°C until further use.

Table 2.1 Antibiotics used for selection of transformant *E. coli.*

Gene [a]	Drugs [b]	Class	Mode of action [c]
Bla/Amp	Ampicillin 100	Beta-lactam	Cell wall synthesis inhibition
Kan	Kanamycin 50	Aminoglycoside	Mis-translation by binding ribosomal 30S
Cam	Chloramphenicol 25	-	Peptidyl translocation inhibition by 50S binding
Tet	Tetracycline 12.5	Tetracycline	Protein synthesis inhibition by 30S binding

[a] Selection gene in the transformed plasmid.
[b] Available commercial antibiotics for selection. Concentrations in μg/mL.
[c] Mode of action in prokaryotes

2.1.8 *Escherichia coli* stocks

Transformant *E. coli* stocks were stored at -80°C. Single isolated colonies grown overnight on agar plates were used to inoculate LB Lennox and grown overnight at 37°C 220 rpm shaking. Cells were then transferred to LB 20% glycerol and frozen at -80°C in cryovials. To start a culture from an *E. coli* stock, the frozen stock was briefly scratched with a tip and plated on a pre-warmed LB agar plate and incubated at 37°C overnight.

2.2 Molecular biology methods

2.2.1. Extraction of *Trypanosoma brucei g*enomic DNA (gDNA)

Exponential growth phase BSF or PCF cells (~10^6 cells) were collected by centrifugation at 1,000 x *g* for 5 minutes at RT, washed in cold PBS and resuspended in 200 μL PBS for gDNA extraction using the QIAGEN DNeasy Blood and Tissue kit, following the manufacturer's protocol for cultured cells. gDNA was eluted in nuclease-free water (NFW) and stored at -20°C until further use. gDNA concentration measured

using a Nanodrop spectrophotometer ND1000 by 260 nm light absorbance (Abs) determination. DNA purity estimated by measuring the ratios Abs 260 nm/Abs 280 nm (for protein contamination, optimal value of 2.0) and Abs 260nm/Abs 230 nm (for phenol contamination, optimal value of 2.0).

2.2.2 Plasmid purification

Bacterial plasmids were extracted and purified from *E. coli* single colonies (clones) after overnight growth in liquid LB at 37°C using the QIAGEN Miniprep Kit (for culture volumes < 3mL) or the QIAGEN Midiprep Kit (for volumes <20 mL) following the manufacturer's protocol. Plasmids eluted in NFW and stored at -20°C until further use. Plasmid concentration and purity were estimated as in section 2.2.1.

2.2.3 DNA purification

Non-genomic DNA such as PCR or digestion products (>100 bp) were purified from liquid reactions using the QIAGEN PCR Purification Kit following the manufacturer's protocol. Alternatively, DNA molecules were purified after agarose gel electrophoresis by excising the DNA gel bands of interest (>100 bp) using a scalpel on a UV-transilluminator and using the QIAGEN Gel Extraction Kit following the manufacturer's protocol. DNA eluted in NFW and stored at -20°C until further use. DNA concentration and purity estimated as in section 2.2.1.

2.2.4 RNA extraction and cDNA synthesis

To extract RNA, several *T. brucei* cell types (cultured and collected from tsetse) or tsetse tissues were collected, washed in cold PBS and immediately snap-frozen in liquid nitrogen and stored at -80°C until purification. RNA extraction was then performed using the GeneJet RNA Purification Kit (Thermo Scientific) following the manufacturer's protocol. 2-mercaptoethanol was used as denaturing agent. RNA was eluted in NFW and immediately processed for DNAse treatment using the TURBO DNA-free kit (Ambion) to remove DNA traces. The rigorous (1 hour at 37°C incubation) treatment was performed following the manufacturer's indications. cDNA

synthesis was immediately run at 52°C for 10 minutes using the SuperScript IV reverse transcriptase (Invitrogen), oligo$(dT)_{20}$ primers and <1μg DNA-free RNA. Whenever possible, the starting RNA amount was set equally to normalise equivalent samples. A negative control was set without reverse transcriptase to ensure specific amplification from cDNA only in further PCR reactions.

2.2.5 Polymerase chain reaction (PCR)

PCR has been extensively used in this thesis to amplify DNA molecules from *T. brucei* gDNA and bacterial plasmids for cloning or genotyping purposes. As a general approach, the Taq DNA polymerase 2X Master Mix (New England Biolabs) was used in 20 μL reactions with 0.2 μM forward and reverse primers and <1 ng DNA template. Reaction volume and primer concentrations were eventually modified for convenience. Temperature conditions were set as follows: initial denaturation at 95°C for 30 seconds (1 cycle), denaturation at 95°C for 30 seconds followed by the annealing step (variable temperature) for 30 seconds, and the extension step at 68°C at variable times according to the Taq speed of 1 minute/kb (30 to 40 cycles), and a final extension step at 68°C for 4 minutes before storing the samples at 4°C or -20°C. Annealing temperatures set 4°C below the lowest primer Tm (calculations described in section 2.2.6).

To perform DNA amplifications for cloning, the high-fidelity Q5 DNA polymerase (New England Biolabs) was used instead in 25 μL reactions with 0.5 μM forward and reverse primers and < 1ng DNA template. Reaction volume and primer concentrations were eventually modified for convenience. Temperature conditions set as follows: initial denaturation at 98°C for 30 seconds (1 cycle), denaturation at 98°C for 10 seconds followed by annealing (variable temperature) for 30 seconds and extension at 72°C for variable time considering a polymerase speed of 30 seconds/kb (30 to 40 cycles), final elongation at 72°C for 2 minutes before storing the samples at 4°C or -20°C. Annealing temperatures set 3°C above the lowest primer Tm (calculations described in section 2.2.6). Primer sequences in **Appendix I**.

2.2.6 Primer design

Primers for PCR, semi-quantitative and quantitative RT-PCR were manually designed and analysed using OligoAnalyzer 3.1 (Integrated DNA Technologies). Primer design aimed for Tm ~60°C calculated under standard buffer conditions (Na^+ 50 mM, Mg^{2+} 1.5 mM, dNTPs 0.2 mM, oligonucleotides 0.4 μM), a length of 18 to 22 bp, GC content between 40% and 60% and to meet the 3' GC clamp rule. Primers were thermodynamically checked to avoid the formation of hairpins and primer-primer interactions at the annealing temperature. To avoid unspecific annealings, primers were then BLASTed against the genome of *T. b. brucei* TREU972 reference strain (TriTrypDB.org) and the *Glossina morsitans morsitans* Yale strain transcriptome GmorY1.8 (VectorBase.org) only if trypanosomes were extracted from tsetse.

For cloning, primers were analysed in the same way described above and restriction targets were manually added to the 5' end preceded by the bases 'CGC' to allow the restriction enzyme to properly settle on the DNA strand. For example, the sequence 5'- CGCAAGCTT- 3' was added upstream the annealing sequence to add the restriction target *HindIII* (underlined) to the resultant amplicon after PCR amplification. These sequences were not considered during Tm calculations.

Primers for qRT-PCR (using the intercalant dye SYBR Green) were designed following the same procedure described above plus considering the length of the amplicon to be between 80 bp to 200 bp for optimal detection and quantification. The resultant Tm and melting curve of the amplicons were predicted using uMelt (Wittwer Lab, University of Utah).

2.2.7 Agarose gel electrophoresis

Products from PCR, RT-PCR and restriction reactions were resolved by gel electrophoresis. Agarose (0.8% to 2%) Tris-Acetate EDTA (TAE) gels stained with the DNA-intercalant dye SYBR Safe (Invitrogen) or ethidium bromide were loaded with DNA samples mixed with loading buffer (contains glycerol, EDTA and colorant) were run in TAE buffer at 90V to 120V for variable times. Resolved gels were imaged using a GelDoc EZ system (Bio-Rad) under variable exposure conditions.

2.2.8 Semi-quantitative RT-PCR

Semi-quantitative RT-PCR was used to determine relative gene expression levels in samples with originally reduced content of RNA (i.e. RNA extracted from SG trypanosomes) or to determine the expression of genes with high sequence identity (i.e. *misp* homologs, Chapter 3), virtually impossible to differentiate by the more restrictive quantitative RT-PCR. PCR reactions were setup using the Taq DNA polymerase as described in section 2.2.5 and using cDNA as template. Reactions were paused and sampled several times between PCR cycles 30 to 40 to analyse the products during the exponential phase of the reaction and not at plateau phase when products saturate. Samples were then resolved by gel electrophoresis as described in section 2.2.7 under identical conditions and band intensities determined using the ImageLab software (Bio-Rad) after imaging. *T. brucei* gene expression was normalized to the expression of *tert* (telomerase reverse transcriptase, Tb927.11.10190; Brenndorfer et al. 2010). Detection of *tert* was obtained using the primers TERT_RT-PCR_F and TERT_RT-PCR_R. Primer sequences in **Appendix I**.

2.2.9 Real-time quantitative RT-PCR (qRT-PCR)

qRT-PCR was the standard method used to quantify gene expression in *T. brucei.* The Luna qPCR Universal 2X Mastermix (SYBR Green and ROX reference dye, New England Biolabs) was used in 20 μL reactions with 0.25 μM forward and reverse primers and 1μL cDNA as template. Samples were run in triplicates in 96-well plates in an Agilent Mx3005P qPCR machine. PCR temperature conditions were set as follows: initial denaturation at 95°C for 1 minute, followed by 40 cycles of 95°C for 15 seconds and 60°C for 30 seconds before measuring fluorescence. Product melting curves were determined with a final dissociation curve from 55°C to 95°C (0.5°C fluorescence measurements). Gene expression was normalised to the expression of *tert* (Brenndorfer and Boshart, 2010) whose detection was made using the primers TERT_qPCR_F and TERT_qPCR_R. Amplification plots and melting curves were analyzed using the MxPro software (Agilent). The threshold was set automatically on the selected samples and the gene expression fold changes were calculated from the resultant cycle threshold values (Ct) applying the $2^{-\Delta\Delta Ct}$ method (Livak et al., 2001).

2.2.10. Restriction of DNA

DNA restriction using restriction enzymes was extensively used in molecular cloning. Restriction of plasmid DNA (<1 μg) was set in 20 μL reactions with FastDigest buffer (Thermo Scientific) and 1 μL of each FastDigest restriction enzyme. Reactions were incubated at 37°C for <30 minutes for restriction (variable time according to manufacturer's protocol) and heat-inactivated. Alternatively, PCR products (<200 ng) were restricted in 30 μL reactions under similar conditions.

2.2.11 Plasmid dephosphorylation

Restricted plasmids had to be eventually dephosphorylated in both 5' and 3' free ends to prevent self-ligation using a recombinant shrimp alkaline phosphatase (rSAP, New England Biolabs). A 20 μL reaction was set with 1 μg restricted plasmid, CutSmart buffer (New England Biolabs) and 1 unit rSAP, incubated at 37°C for 30 minutes and heat-inactivated at 65°C for 5 minutes. The dephosphorylated plasmids were then used in further ligation reactions (section 2.2.12).

2.2.12. DNA ligation

DNA ligation reactions were extensively used in molecular cloning. Typically, a restricted plasmid was ligated to a smaller dsDNA molecule called insert (e.g. PCR product) restricted with the same restriction enzymes or different ones yielding compatible ends. Ligation reactions were set in 20 μL containing 50 ng restricted plasmid, a variable amount of restricted insert (see **Figure 2.3**) and 400 units T4 DNA ligase with the appropriate buffer (New England Biolabs). Reactions were incubated either at RT for 2 hours or at 16°C overnight and posteriorly heat-inactivated at 65°C for 20 minutes. Ligation reactions were kept at 4°C until further use.

$$\text{insert amount (ng)} = \frac{\text{plasmid amount (50 ng)} \cdot \text{insert size (kb)}}{\text{plasmid size (kb)}} \cdot \frac{\text{insert molecules (3)}}{\text{plasmid molecules (1)}}$$

Figure 2.3. Equation for the calculation of the amount of insert to use in a typical ligation reaction. By default, 50 ng plasmid are used, and the 3:1 ratio (insert:plasmid) can be increased up to 8:1 in challenging ligation reactions.

2.2.13. *Escherichia coli* transformation

Ligation reaction products were transformed into *E. coli* strain DH5α competent cells ($CaCl_2$-acquired competency, Invitrogen). Cells were removed from -80°C and thawed at 0°C before adding ~10 ng ligation reaction and incubated on ice for 30 minutes. Cells were then heat-shocked at 42°C for 20 seconds, incubated on ice for 2 minutes and SOC medium was added for recovery at 37°C for 1 hour, 220 rpm shaking. Cells were concentrated by centrifugation at 6,000 x g for 3 minutes and plated on LB agar plates containing the selection antibiotic (see **Table 2.1**). Plates were incubated at 37°C overnight. Isolated bacteria colonies (clones) were then streaked on a new agar plate for further use. Transformation of *E. coli* Rosetta 2 (DE3) pLysS and Rosetta-gami 2 (DE3) pLysS strains was done as described above except for a first incubation on ice for 5 minutes and a heat-shock at 42°C for 30 seconds. Both strains were plated on LB agar plates supplemented with chloramphenicol and tetracycline (to maintain parental markers) besides the selection antibiotic.

2.2.14. Plasmid linearisation

The plasmids containing the final DNA constructs (entire plasmid or section) for *T. brucei* transfection were linearised to increase the efficiency of homologous recombination. To linearise 20 μg of plasmid requiring restriction at the *NotI* site, 4 reactions of 50 μL were setup containing each 5 μg plasmid, 1 unit NotI high-fidelity restriction enzyme (New England Biolabs) and the appropriate buffer (CutSmart) and incubated at 37°C for 16 hours before heat-inactivating the enzyme at 65°C for 20 minutes. Plasmids requiring partial linearisation at *HindIII* and *SacI* sites were restricted using 1 unit of FastDigest *HindIII* and *SacI* (Thermo Scientific) in 50 μL

reactions with 5 μg plasmid and FastDigest buffer, incubated at 37°C for 4 hours and heat-inactivated at 80°C for 10 minutes. Linearised plasmids were kept at -20 °C until further use.

2.2.15. DNA precipitation

Linearised plasmids were precipitated before transfection to purify, sterilise and concentrate the DNA. 20 μg plasmid was incubated in ice-cold 300 mM sodium acetate pH=5, 70% ethanol (-20°C) and 57 μg/mL glycogen for 10 minutes and spun down at 13,000 x *g* for 15 minutes at 4°C to pellet the DNA. The pellet was washed in 70% ethanol and air-dried before resuspension in ~12 μL NFW (for a concentration > 1 μg/μL). DNA was stored at -20°C until further use in parasite transfections (sections 2.2.19 and 2.2.20).

2.2.16 General cloning procedure

To create DNA constructs to either express recombinant proteins in *E. coli* or to transfect into *T. brucei* for the generation of transgenic cell lines (sections 2.2.17 and 2.2.18), a series of cycles of cloning procedures were performed as follows. DNA coding sequences were typically amplified by PCR (high-fidelity polymerase) (section 2.2.5), resolved by gel electrophoresis (section 2.2.7) to identify the desired DNA band based on relative molecular size, and purified from the agarose gel (section 2.2.3). Alternatively, DNA coding sequences contained in plasmids were obtained by restriction of the surrounding restriction targets and purified from resolved gels (section 2.2.10). DNA coding sequences were then ligated (section 2.2.12) into pre-restricted plasmids with the same restriction targets or others yielding compatible ends. Ligation products were transformed into *E. coli* DH5α (section 2.2.13) for selection of clonal transformants. The resultant plasmids were purified from bacteria cultures (section 2.2.2) and either used in subsequent cloning rounds or in different downstream procedures. These cloning methods are described here as a general approach and they were eventually modified in particular cases. Specific cloning procedures for the creation of plasmids for the expression of recombinant proteins in

E. coli detailed in section 2.2.17, and for the generation of *T. brucei* transgenic cell lines described in section 2.2.18.

2.2.17. Creation of plasmids for the expression of MISP recombinant proteins

For the expression of recombinant MISP380 in *E. coli* (**Figure 3.12-A**), the coding sequence of *misp380* lacking the signal and GPI anchor peptide sequences was amplified by PCR from gDNA extracted from *T. b. brucei* PCF using the primers MISP_-SP_F_NcoI and MISP_-GPI_R_XhoI, containing the restriction sites *NcoI* and *XhoI* (**Appendix I**). Briefly, the plasmid pET28a (Novagen) and PCR product were digested with *NcoI* and *XhoI* and ligated for the insertion of the amplicon into the plasmid in frame fused to a poly-histidine tag (6xHis) at C-terminus. The ligation reaction was transformed into *E. coli* DH5α for selection and the sequence was corroborated by Sanger sequencing. The pET28-MISP380-6xHis plasmid was then transformed into *E. coli* Rosetta 2 (DE3) pLysS for protein expression.

For the expression of recombinant MISP360 (**Figure 3.14**), the coding sequence of *misp360* (lacking both the signal and GPI anchor peptide sequences) was obtained by PCR as described above using the primers MISP_-SP_F_NdeI and MISP_-GPI_XbaI, containing the restriction sites *NdeI* and *XbaI* (**Appendix I**), for cloning into the plasmid pCOLD II (Takara) in frame fused to a 6xHis-tag at N-terminus. Alternatively, for the recombinant expression of either the N-terminus or C-terminus domains of MISP360, the corresponding coding sequences were amplified by PCR using the primers MISP_-SP_F_NdeI and MISP_N_R_XbaI, and MISP_C_F_NdeI and MISP_-GPI_R_XbaI, respectively. Both sequences contained the restriction sites *NdeI* and *XbaI* (underlined) for cloning into the plasmid pCOLD II fused to a N-terminus 6xHis-tag. The resultant plasmids pCOLD-6xHis-MISP360N+C, pCOLD-6xHis-MISP360N and pCOLD-6xHis-MISP360C were checked by sequencing as described above and then transformed into *E. coli* Rosetta-gami 2 (DE3) pLysS for protein expression.

2.2.18 Creation of plasmids for the generation of *T. b. brucei* mutant lines

2.2.18.1 Generation of plasmid pDEX577-HA/eGFP-MISP360

The plasmid pDEX577-HA/eGFP-MISP360 was built for the tetracycline-inducible overexpression of chimeric MISP360 (Tb927.7.360), which has an HA epitope and an enhanced GFP (eGFP) fused at N-terminus, downstream the MISP signal peptide (**Figure 3.24-A**). First, the HA coding sequence was obtained by annealing (95°C for 5 minutes and progressive cooling at RT) of the complementary oligonucleotides HA_F_HindIII and HA_R_SpeI and restriction of the resultant DNA sequence with *HindIII* and *SpeI*. This fragment was then ligated in frame into the plasmid pDEX577 (Kelly et al. 2007, **Table 2.2**) using the same restriction sites. Second, the coding sequence of the MISP360 signal peptide was obtained by annealing of the complementary oligonucleotides MISP360_SP_F_HindIII and MISP360_SP_R_HindIII as described above; the DNA fragment was then restricted with *HindIII* and ligated into the pDEX577-HA plasmid previously generated. Insertion of the signal peptide sequence in the correct direction was corroborated by sequencing. Lastly, the coding sequence for MISP360 without signal peptide was amplified by PCR from *T. brucei* gDNA using primers MISP360_CDS-SP_F_XbaI and MISP360_CDS_R_BglII, containing the restriction targets *XbaI* and *BglII*. Plasmid pDEX577, containing the signal peptide, HA and eGFP was restricted with the enzymes *XbaI* and *BamHI* and ligated with the PCR product previously digested with *XbaI* and *BglII* (*BamHI* and *BglII* yield compatible sticky ends). The sequence of the resultant plasmid (pDEX577-HA/eGFP-MISP360) was checked by sequencing at the insertion site.

2.2.18.2 Generation of pALC14 plasmids for the RNAi knockdown of *T. b. brucei misp, rbp6* and *stt3* genes

Three different plasmids were made for the tetracycline-inducible overexpression of a short-hairpin dsRNA to RNAi silence the expression of all *misp* (e.g. Tb927.7.360), *rbp6* (Tb927.3.2930) and *stt3* (e.g. Tb927.5.890) transcripts (**Figures 3.33-A, 4.8-A and 5.4-A**). For the knockdown of *misp,* a unique short 491 bp coding sequence of *misp* fully conserved between the five homologs was amplified by PCR in both sense and antisense directions using the primer pairs

MISP_RNAi_SENSE_F_XhoI and MISP_RNAi_SENSE_R_BamHI (containing the targets *XhoI* and *BamHI*) and MISP_RNAi_ANTI_F_NdeI and MISP_RNAi_ANTI_R_HindIII (containing the targets *NdeI* and *HindIII*). Both sense and antisense fragments were inserted into the plasmid pALC14 (MacGregor et al., 2013) (**Table 2.2**) using the restriction sites *HindIII, NdeI, XhoI* and *BamHI* surrounding the 460 bp stem loop sequence. The inserted sequences were checked by sequencing of the resultant plasmid pALC14-MISP-RNAi.

For the knockdown of *stt3,* a conserved 537 bp coding sequence of *stt3* in both sense and antisense directions was amplified by PCR using the primers STT3_RNAi_SENSE_F_XhoI and STT3_RNAi_SENSE_R_BamHI, and STT3_RNAi_ANTI_F_NdeI and STT3_RNAi_ANTI_R_HindIII , and inserted into the plasmid pALC14 under identical restriction sites and the resultant plasmid pALC14-STT3-RNAi was checked by sequencing.

For the knockdown of *rbp6*, a 516 bp coding sequence of *rbp6* in sense and antisense directions was amplified by PCR using the primers RBP6_RNAi_SENSE_F_XhoI and RBP6_RNAi_SENSE_R_BamHI, and RBP6_RNAi_ANTI_F_NdeI and RBP6_RNAi_ANTI_R_HindIII, and was inserted into pALC14 using the same restriction sites as described above. The resultant plasmid pALC14-RBP6-RNAi was checked by sequencing.

2.2.18.3 Generation of plasmids for the overexpression of RBP6

The plasmids pMS-3xHA/sfGFP-RBP6 and pDEX577-HA-RBP6 were made for the tetracycline-inducible overexpression of RBP6 (Tb927.3.2930) either fused to a tandem of three HA epitopes and superfolder GFP (sfGFP) or a single HA epitope only, respectively, both tagged at N-terminus (**Figure 4.1**). For the first construct, the coding sequence of *rbp6* was amplified by PCR from *T. brucei* gDNA using the primers RBP6_CDS_F_XhoI and RBP6_CDS_R_BamHI containing the restriction sites *XhoI* and *BamHI* and inserted into the pMS plasmid (Serricchio et al., 2013), **Table 2.2**) under the same restriction targets. The sequence coding for three tandem HA epitopes followed by the sfGFP was obtained by PCR from the plasmid pSIG (Gadelha et al., 2015) using the primers 3xHA_sfGFP_F_HindIII and 3xHA_sfGFP_R_XhoI containing the restriction sites *HindIII* and *XhoI* and inserted

into the plasmid pMS-RBP6 into the same restriction sites. The resultant plasmid pMS-3xHA/sfGFP-RBP6 was checked by sequencing.

To create the second plasmid, the coding sequence of *rbp6* was amplified by PCR from *T. b. brucei* gDNA using the primers RBP6_HA_CDS_F_HindIII (containing the HA coding sequence) and RBP6_CDS_BamHI with the restriction targets *HindIII* and *BamHI* and ligated into the plasmid pDEX577 using the same restriction targets. The resulting plasmid pDEX577-HA-RBP6 was corroborated by sequencing of the insert.

2.2.18.4 Generation of plasmids for the overexpression of MYND, GDH and FT

The plasmid pMS (**Table 2.2**) (Serricchio and Butikofer, 2013) was used to overexpress (tetracycline-inducible) the hypothetical protein MYND (Tb927.4.4670), the glutamate dehydrogenase (GDH) (Tb927.9.5900) and one folate transporter (FT) homolog (Tb927.8.3630) (**Figure 4.17**). The coding sequences were obtained by PCR amplification from *T. brucei* gDNA. To amplify *mynd,* the primers MYND_CDS_F_HindIII and MYND_CDS_R_XbaI *(HindIII* and *XbaI* targets) were used, for *gdh* the primers GDH_CDS_F_XbaI and GDH_CDS_R_XbaI (*XbaI* only; insert direction corroborated by sequencing) were used, and for the folate transporter the primers FT_HA_CDS_F_HindIII (containing the HA coding sequence) and FT_CDS_R_BamHI (*HindIII* and *BamHI* targets) were used. The coding sequences of *mynd* and *gdh* were inserted under their corresponding restriction sites into the plasmid pMS containing a previously-inserted single HA epitope inserted under the restriction sites *XbaI* and *BamHI* at C-terminus. On the other hand, the coding sequence of the folate transporter was inserted into the pDEX577 (Kelly et al., 2007) under the restriction sites *HindIII* and *BamHI* as the amplicon contained the HA sequence at N-terminus. The resulting pMS-MYND-HA, pMS-GDH-HA and pDEX577-HA-FT plasmids were checked by sequencing of the inserts.

2.2.18.5 Generation of plasmids for the creation of *stt3* and *rft1* conditional-null mutants

For the creation of the *stt3* and *rft1* conditional-null mutants, several DNA constructs were built (**Figure 5.2**). For the knockout of the *stt3* array (Tb927.5.880, Tb927.5.890, Tb927.5.900 and Tb927.5.910), flanking regions at 5' and 3' of the array (556 bp and 570 bp respectively) were obtained by PCR amplification from *T. brucei* gDNA using the primer pairs STT3_5_F_HindIII and STT3_5_R_EcoRI (*HindIII* and *EcoRI*), and STT3_3_F_XbaI and STT3_3_R_SacI (*XbaI* and *SacI*). These regions were inserted using the same restriction sites into pKO plasmids (modified pBlueScript, Agilent), either containing the resistance marker gene *puro* (puromycin) or *blast* (blasticidin), flanking the marker genes. For knocking out *rft1* (Tb927.11670), the same procedure was applied with a 5' flanking region of 397 bp and a 3' region of 427 bp using the primer pairs RFT1_5_F_HindIII and RFT1_5_R_EcoRI (*HindIII* and *EcoRI*), and RFT1_3_F_XbaI and RFT1_3_R_SacI (*XbaI* and *SacI*).

The *rft1* addback construct was made by inserting the coding sequence of *rft1* obtained by PCR with the primers RFT1_CDS_F_HindIII and RFT1_CDS_R_XhoI (*HindIII* and *XhoI*) into the plasmid pMS with a single HA epitope sequence inserted at C-terminus under the targets *XhoI* and *BamHI*.

2.2.18.6. Generation of plasmid pALC14-mNeonGreen

The plasmid pALC14-mNeonGreen was built for the tetracycline-inducible overexpression of the reporter mNeonGreen fluorescent protein in BSF. The coding sequence of *mNeonGreen* was obtained by PCR amplification using the primers mNG_CDS_F_HindIII and mNG_CDS_R_BamHI containing the restriction sites *HindIII* and *BamHI*. The DNA fragment was ligated into the plasmid pALC14 (**Table 2.2**) (MacGregor et al., 2013) digested with the same restriction enzymes, resulting into the plasmid pALC14-mNeonGreen, which was sequenced before use.

Table 2.2. Plasmids used for the creation of *T. b. brucei* transgenics.

Plasmid	Promoter [(a)]	*E. coli* [(b)]	*T. brucei* [(c)]	Insertion site [(d)]
pDEX577	Procyclin (EP)	Ampicillin	Phleomycin/ Zeocin	177 bp repeats
pALC14	Procyclin (EP)	Ampicillin	Puromycin	rDNA spacer
pMS	rDNA	Ampicillin	Phleomycin/ Zeocin	rDNA spacer

[(a)] Promoter driving the expression of the transgene in *T. brucei*
[(b)] Drug selection in *E. coli*
[(c)] Drug selection in *T. brucei*
[(d)] Recombination site in *T. brucei*

2.2.19 Transfection of *Trypanosoma brucei brucei* BSFs

T. b. brucei strain AnTat 1.1 90:13 BSF cells were used for transfection at culture densities between 4 x 10^5 cells/mL and 9 x 10^5 cells/mL (exponential growth phase). Cells were spun down at 1,000 x *g* for 10 minutes at 4°C and washed in Tb-BSF ice-cold buffer (90 mM $NaPO_4$ buffer, 5 mM KCl, 0.15 mM $CaCl_2$, 50 mM HEPES, pH=7.3). Per transfection, 4 x 10^7 cells were mixed with 10 μg linearised DNA construct in 100 μL Tb-BSF buffer and electroporated in a cuvette using the Amaxa 4D Nucleofector X unit, program FI-115. Cells were then immediately transferred to a flask containing 10 mL of pre-warmed (37°C) HMI-9 medium, supplemented with 10% FBS, G418 (2.5 μg/mL) and hygromycin B (4 μg/mL) and let recover for 6 hours at 37°C and 5% CO_2. After recovery, the selection drug (see **Table 2.3** for concentration details) was added and serial dilutions in 24-well non-treated plates were performed for further selection of clonal cell lines. The following controls were added: 1) transfection negative control (no DNA added) and 2) drug selection control (parental cell line with and without selection drug). After 5 days, transfectant cell lines were typically screened based on the corresponding phenotype and the transgene insertion was checked by PCR.

2.2.20 Transfection of *Trypanosoma brucei brucei* PCF

T. b. brucei strain AnTat 1.1 90:13 or strain Lister 427 29:13 PCF cells were used for transfection at exponential growth phase at densities between 5 x 10^6 cells/mL and 10^7 cells/mL. Cells were spun down at 1,000 x *g* for 5 minutes at RT and washed in Tb-BSF buffer. Per transfection, 4 x 10^7 cells were mixed with 10 μg linearised DNA construct in 100 μL Tb-BSF buffer and electroporated in a cuvette using the Amaxa 4D Nucleofector X- unit, program FI-115. Cells were then immediately transferred to a flask containing 10 mL of pre-warmed (27°C) SDM-79 medium, supplemented with 15% FBS, G418 (15 μg/mL) and hygromycin B (50 μg/mL) and let recover for 2 hours at 27°C and 5% CO_2. After recovery, serial dilutions were done in SDM-79 15% FBS and 20% conditioned medium (0.22 μm-filtered supernatant from parental cell line at exponential phase of growth) in 96-well plates for further selection of clonal cell lines, and cells were incubated overnight at 27°C and 5% CO_2. 24 hours post-transfection, the selection drugs were added (see **Table 2.3** for concentration details). The controls described in section 2.2.19 were also applied to PCF transfections. After 14 days, transfectant cell lines were typically screened based on the corresponding phenotype and the transgene insertion was checked by PCR. Alternatively, the entire procedure of PCF transfection was carried out using DTM medium instead of SDM-79 under identical conditions for the transfection of DTM-adapted PCF.

Table 2.3 Drugs used for the selection of transfectant *T. b. brucei* BSF and PCF.

Gene [a]	Drugs [b]	BSF selection [c]	PCF selection [c]	Mode of Action [d]
Ble	Phleomycin/ Zeocin	Phleomycin 2.5 + Zeocin 5	Zeocin 10	DNA binding and cleavage
Puro	Puromycin	Puromycin 1	Puromycin 10	Inhibits peptidyl transfer
Blast	Blasticidin	Blasticidin	Blasticidin	Blocks protein synthesis
*Neo**	Neomycin/ G418	G418 2.5	G418 15	Blocks protein synthesis
*Hyg**	Hygromycin B	Hygromycin B 4	Hygromycin B 50	Blocks protein synthesis

[a] Selection marker gene for the transgene insertion.
**Neo* and **Hyg* not used for selection but for maintenance of the parental markers.
[b] Available commercial drugs for selection.
[c] Concentration in μg/mL.
[d] Affects both eukaryotes and prokaryotes

2.3 Methods for protein analysis

2.3.1 Expression of recombinant proteins in *E. coli*

For the expression of MISP recombinant proteins, several plasmids were created and transformed into *E. coli* BL21 (DE3) strains (section 2.2.17) for the inducible overexpression of the proteins. The selected clones of *E. coli* Rosetta (DE3) pLysS or Rosetta-gami 2 (DE3) pLysS were set to exponential phase growth in ~50 mL Lennox LB supplemented with kanamycin (pET28) or ampicillin (pCOLDII), chloramphenicol and tetracycline (**Table 2.1**), and 40 mg/L L-leucine at 37°C 220 rpm shaking for 2 hours until an optical density of OD_{600nm}=1.2 was reached. These cells were then used to inoculate 4 litres LB with antibiotics and incubated at 37°C 220 rpm shaking until a OD_{600nm}= 0.5 to 0.7 was reached. Protein overexpression was then induced with 0.5 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) and cultures were incubated at 16°C overnight reaching OD_{600nm} = ~1.2. Cultures were set at 4°C to stop growth before harvesting.

2.3.2 Ni-NTA chromatography

E. coli cells from section 2.3.1 were harvested by centrifugation in 250 mL bottles at 3,500 x *g* for 30 minutes at 4°C. The supernatant was discarded, the bacteria pellet transferred to a smaller tube and spun down for 20 more minutes at 4°C. Medium leftovers were removed and pellet was frozen and kept at -80°C at least 24 hours before proceeding with protein purification. Pellets were dissolved in lysis buffer (50 mM imidazole, 20 mM HEPES pH=7.8, 100 mM NaCl) supplemented with EDTA-free protease inhibitors (Roche) and 10 mg/mL DNAse I (Sigma), and sonicated on ice at 70% amplitude for 15 pulses of 10 seconds each. The solution was centrifuged at 35,000 x *g* for 30 minutes at 4°C, the soluble fraction (containing the protein) was collected and filtered using a 0,22 μm filter. This was then incubated in batch with 2mL Ni-NTA HisPur Superflow (Thermo Scientific) agarose (equilibrated with 10 volumes of lysis buffer) for 3 hours at 4°C and posteriorly assembled in a PD-10 column for several washes with washing buffer (20 mM HEPES, 100 mM NaCl). Lastly, two 0.5 mL elutions in elution buffer 1 (100 mM imidazole washing buffer) followed by other 13 elutions of 0.5 mL of elution buffer 2 (300 mM imidazole washing buffer) were collected. Protein concentrations were estimated using a Nanodrop ND1000 spectrophotometer from the absorbance at 280 nm corrected with the molar extinction coefficient of the corresponding recombinant protein. The resulting flow-through, washes and elution samples were resolved in SDS-PAGE gels for purity assessment.

2.3.3 Size exclusion chromatography (SEC)

The elutions with the apparent purest content of recombinant protein from section 2.3.2 were pooled and further purified by SEC in an AKTA Fast Protein Liquid Chromatography (FPLC) system (GE Healthcare) using a Superdex 200 10/30 GL analytical column (GE Healthcare). The column was first washed with degassed dH_2O and then equilibrated with degassed buffer (20 mM HEPES, 100 mM NaCl) before running the standard molecular size markers (Sigma). After several washes with buffer, a Ni-NTA-purified MISP380 pooled sample of ~0.5 mL was run into the column at ~1.5 MPa pressure, and 35 fractions (of 0.5 mL each) were collected and analysed by SDS-PAGE.

2.3.4 Production of anti-MISP polyclonal antibodies

The polyclonal anti-MISP was obtained through immunisation of a single rabbit (Davids Biotechnologie GmbH) with 0.3 mg of SEC-purified MISP380 recombinantly expressed in *E. coli* Rosetta 2 (DE3) cells (sections 2.3.1, 2.3.2, 2.3.3). The rabbit was bled at 63 days post immunisation and the anti-sera was processed for IgG purification through a protein A column. Purified IgGs were run into a column containing recombinant MISP380 for affinity-purification of specific IgGs with affinity for MISP380. The anti-MISP380 IgGs were passed through a column containing *E. coli* Rosetta 2 (DE3) cell lysate to deplete IgGs with affinity for bacterial proteins. The resultant purified anti-MISP380 IgG sample was tested by immunoblotting and found to remain stable after two freeze-thaw cycles. Antibodies were kept at -20°C for long term storing or at 4°C for frequent use. Non-purified anti-sera and pre-immune sera were obtained and used in control reactions.

2.3.5. SDS-PAGE

Variable amounts of different *T. brucei* stages or purified proteins were heat-denatured at 98°C for 8 minutes in Laemmli sample buffer (Laemmli, 1970) containing DTT as reducing agent. Samples were resolved in 12.5% SDS-PAGE Bis-Tris gels at 50V for 20 minutes followed by ~2 hours at 120V in running buffer (0.1% SDS, 25 mM Tris, 192 mM glycine, pH=8.3) using a Bio-Rad mini-Protean system. Gels were either stained with InstaBlue UltraFast (Expedeon) to reveal protein bands or further used for immunoblotting or lectin blotting. Stained gels were imaged using a GelDoc EZ system and ImageLab software (Bio-Rad).

2.3.6 Western blotting

Proteins resolved in SDS-PAGE gels were transferred to a polyvinylidene fluoride (PVDF) membrane pre-activated in 100% methanol for 30 seconds and equilibrated in transfer buffer (20% methanol, 25 mM Tris, 192 mM glycine, pH=8.3) for 15 minutes. Wet transference was normally carried out at 90V for 1 hour in cold conditions. Membranes were briefly washed in washing buffer (PBS-Tween 0.1 % v/v) and Ponceaux S-stained for rapid visualisation of the transferred proteins. After

washing off the stain, membranes were blocked in blocking buffer (PBS-Tween 0.1 % v/v, 5 % w/v skim milk powder) for 1 hour at RT. Membranes were then probed with the primary antibody (**Table 2.4**) in blocking buffer overnight at 4°C or 1 hour at RT. Membranes were extensively washed in washing buffer and secondary antibodies (**Table 2.4**) conjugated to horse radish peroxidase (HRP) were incubated in blocking buffer for 1 hour at RT. After extensive washings in washing buffer, membranes were probed with Super Signal West Dura or Femto Dura luminol substrate (Thermo Scientific) for 5 minutes in the dark and exposed to X-ray Amersham Hyperfilm (GE Healthcare) at variable exposition times. After exposure, membranes were stained in 0.2% nigrosin for 15 minutes for sample loading control visualisation.

Table 2.4. Antibodies used in western blotting.

Antibody	Antigen	Type	Host	Dilution
α-*Sodalis*	*Sodalis* Hsp60	Monoclonal hybridoma sup.	Mouse	1:10
α-MISP	*T. brucei* MISP (all)	Polyclonal purified	Rabbit	1:2,000
α-BARP	*T. brucei* BARP (all)	Polyclonal purified	Rabbit	1:2,000
α-CRD	*T. brucei* CRD epitope	Polyclonal purified	Rabbit	1:500
α-HA	HA epitope	Polyclonal purified	Mouse	1:10,000
α-EP	*T. brucei* EP procyclin	Monoclonal purified	Mouse	1:2,000
α-rabbit-HRP	Rabbit IgG (H+L)	Polyclonal purified	Goat	1:20,000
α-mouse-HRP	Mouse IgG (H+L)	Polyclonal purified	Goat	1:20,000

2.3.7 Lectin blotting

PCF procyclins resolved in SDS-PAGE were transferred at 90V for 1 hour to nitrocellulose membranes pre-equilibrated in transfer buffer for 15 minutes. Lectin specificity control samples were either digested (or mock treated) under denaturing conditions with 25 units/μL PNGase F (New England Biolabs) at 37°C overnight to deglycosylate *N*-glycoproteins. Membranes were briefly washed in washing buffer and blocked in PBS-Tween 0.05 % v/v, BSA 1% w/v overnight at 4°C. Biotinylated concanavalin A (Vector Laboratories) was incubated at 1 μg/mL for 30 minutes at RT and washed off extensively in washing buffer. Lectin control samples were instead incubated with pre-saturated biotinylated ConA (ConA pre-incubated with 200 mM α-methyl-D-mannoside and 200 mM α-methyl-D-glucoside at 27°C for 30 minutes). Membranes were incubated with streptavidin-HRP (1:100,000 dilution) (Vector Laboratories) for 1 hour at RT. After washing, membranes were probed with Super Signal Pico Plus (Thermo Scientific) for 5 minutes in the dark and developed and stained with nigrosin as described in section 2.3.5.

2.3.8 Procyclin extraction from *T. brucei* PCF cells

PCF cells were harvested, washed in PBS and pelleted by centrifugation at 2,000 x *g* for 5 minutes. Cell pellets were snap-frozen in liquid nitrogen and lyophilised. Lyophilised cells were ground using a pestle and dissolved in 150 μL chloroform-methanol-water (10:10:3, v/v/v) solution for lipid extraction. After sonication in an ultrasonic bath, samples were spun down at 6,000 x *g* for 3 minutes at RT and the supernatant was discarded. This process was repeated 3 times before the pellets were dried in the SpeedVac concentrator. 100 μL 9% 1-butanol were added to the pellets, and the samples were sonicated and centrifuged at 10,000 x *g* for 3 minutes at RT. The supernatant, rich in procyclins, was collected. This processed was repeated 3 times, and the supernatants were pooled, dried in a SpeedVac and stored at -80°C until further use.

2.3.9 nLC-MS/MS Proteomics

Samples from naïve and *T. brucei*-infected tsetse saliva were lyophilised and sent to the Dundee University Fingerprints Proteomics Facility for in-solution trypsin digests and mass spectrometry. Following trypsinisation, the resulting peptides were analyzed by nano-liquid chromatography tandem mass spectrometry (nLC-MS/MS) using the Ultimate 3000 RSLC nano coupled to LTQ-Orbitrap Velos Pro MS. Resulting spectra from HR-MS/MS analysis of *T. b. brucei* and *G. m. morsitans* samples were searched with Proteome Discoverer (PD) 2.1.1.21 (Thermo Scientific) and filtered with an estimated false-discovery rate (FDR) of 1%. The PD settings were as follows: HCD MS/MS; cysteine carboxyamidomethylation as fixed modification; methionine oxidation and acetylation on any amino acid as variable modifications; fully tryptic peptides only; up to 2 missed cleavages; parent ion mass tolerance of 10 ppm (monoisotopic); and fragment mass tolerance of 0.6 Da (in Sequest) and 0.02 Da (in PD 2.1.1.21) (monoisotopic). Tandem MS/MS spectra were searched against the following databases: a combined protein database of *T. b. brucei, T. b. gambiense* and *T. b. rhodesiense* (18,729 entries from TriTrypDB, October 2017) and (25,450 entries from UniProtKB, October 2017); and *G. m. morsitans* (12,494 entries from VectorBase, May 2017). A database of common contaminant sequences (MaxQuant) was also added to the combined database. The resulting PD dataset was further processed through Scaffold Q+ 4.8.2 and perSPECtives (Proteome Software). A protein threshold of 95%, peptide threshold of 95%, and a minimum number of 1 peptide were used for identification of proteins. Manual inspection of the MS/MS spectra was always performed for the identification of any protein based on a single peptide. Using BLAST2GO (BioBam), proteins were BLAST searched against the non-redundant NCBI protein database, mapped and categorised according to gene ontology (GO) terms. The *Sodalis glossinidius* proteins were identified using the Liverpool School of Tropical Medicine in house MASCOT MS/MS ion search using a protein database from UniProtKB generated from the latest re-annotated coding sequences in *S. glossinidius* (Belda et al., 2010), and the following settings: 2 missed trypsin cleavages, same peptide modifications described above, peptide tolerance of 1.2 Da, MS/MS tolerance of 1.2 Da and peptide charges set to +1, +2 and +3.

2.3.10 Protein pulldown assay

Streptavidin magnetic beads were washed in TBST buffer and mixed with ~10 μg of recombinant MISP400 or HpHbR (*T. brucei* haptoglobin haemoglobin receptor) (Dr. Olivia MacLeod) biotinylated at C-terminus in TBST, and incubated for 1 hour at RT. Tubes were then put on magnetic racks to trap the beads bound to the recombinant proteins and washed in TBST buffer. Saliva samples from 46 naïve or 25 *T. brucei*-infected tsetse were incubated with the recombinant proteins for 1 hour at RT before using the magnetic rack to perform several washes. All flow-through, washes and bead samples were collected, mixed with Laemmli buffer (Laemmli, 1970) and heated at 98°C for 5 minutes before loaded into 17.5% SDS-PAGE gels. Gels were run in duplicate for both Coomassie Blue and silver staining.

2.4 Tsetse methods

2.4.1 Tsetse fly infections and maintenance

Glossina morsitans morsitans were obtained from the tsetse insectary at the Liverpool School of Tropical Medicine, where they are maintained at 26 ˚C (+/-1 ˚C) and 65-75 % relative humidity. They are fed for 10 minutes every two days on sterile, defibrinated horse blood (TCS Biosciences) through a silicone membrane system. Teneral (unfed, recently emerged) male flies, no older than 48 h, were fed with animal infected blood by combining ~0.5 ml of *T. b. brucei*-infected rat or mouse blood (stored at -195.79°C) with ~5 ml fresh, sterile, defibrinated horse blood for 10-20 minutes. BSF cells were counted and the percentage of 'stumpy' forms estimated based on cell morphology. If flies were infected with cultured BSF, cells were washed in drug-free HMI-9 10% FBS medium three times and mixed with defibrinated horse blood at a standard density of 10^6 cells/mL (variable) before feeding. Alternatively, teneral flies were infected with *T. b. brucei* PCF cells which were washed 3 times in drug-free SDM-79 10% FBS medium by centrifugation at 1,000 x *g* for 5 minutes at RT. Sterile defibrinated horse blood was washed 3 times with the same culture medium by centrifugation at 1,000 x *g* for 10 minutes at RT to remove the complement since PCF are susceptible to it. PCF were mixed with washed blood at a standard density of $5x10^5$ cells/ml and fed to the tsetse for 10-20 minutes. For tsetse infections intended for *T. brucei* MCF collection (section 5.2.3), the bloodmeal was supplemented with 10

mM glutathione to boost MG and PV infections. Uninfected control flies were fed on the same blood type without trypanosomes for the same amount of time. After <24 hours, flies were briefly chilled at 4-8°C to remove the ones that did not engorge. Experimental flies were maintained for up to 4 weeks, feeding on normal defibrinated horse blood every 2-3 days. Alternatively, flies infected with mutant trypanosomes requiring tetracycline activation of the transgene, were fed with normal defibrinated horse blood supplemented with fresh 2 μg/mL doxycycline (dissolved in 70% ethanol). Bloodmeals for control flies were supplemented with the same volume of 70% ethanol only.

2.4.2 Tsetse dissections and *T. brucei* collection

Flies were dissected at variable time points after infection, as late as 30 days post infection and 72 hours after receiving the last bloodmeal. Typically, SGs were first dissected in sterile ice-cold PBS. SG infections were scored by visualising MCF exiting the glands right after dissection or attached EMF in the infected SG tissue. The PV was then dissected on ice-cold PBS with a sharp cut at the junction with the anterior midgut to prevent cross-contamination with MG trypanosomes. PV infections were scored either by squashing the tissue with a coverslip to make it visible or by tearing the organ apart to allow the mesocyclic trypomastigotes to come out. Lastly, MGs were dissected out the abdomen and shred on ice-cold PBS to score the infection. Parasites found in MG, PV and SG were eventually collected by pipetting in ice-cold PBS, washed in PBS by centrifugation at 2,000 x *g* for 5 minutes at 4°C and used alive, fixed or snap-frozen for downstream applications.

2.5 Cell biology methods

2.5.1 Immunofluorescence assay

Cultured or tsetse-extracted *T. b. brucei* cells were washed in PBS and fixed in fresh 4% paraformaldehyde (PFA) for 30 minutes at room temperature (RT). Cells were allowed to settle onto poly-L-lysine coated slides for 20 minutes in a humidity chamber. Alternatively, cells were first air-dried on a poly-L-lysine slide, fixed in 100%

methanol for 10 minutes at -20°C and rehydrated in PBS. PFA-fixed cells were then either permeabilised or not in 0.1% Triton X-100 for 10 minutes at RT. Methanol-fixed cells did not need permeabilisation. Cells were then blocked in 20% FBS vPBS (Voorheis's modified PBS: 138.9 mM NaCl, 2.68 mM KCl, 16 mM Na_2PO_4, 3 mM KH_2PO_4, 45.9 mM sucrose and 10 mM glucose, pH=7.4) for 1 hour, washed in PBS and incubated for 1 hour at RT with the primary antibody in blocking solution (see **Table 2.5** for antibody dilutions). After washing, cells were incubated for 1 hour at RT with the secondary antibody in blocking solution (**Table 2.5**). Cells were then incubated in 300 ng/mL 4',6-diamidino-2-phenylindole (DAPI) (ThermoFisher) for 10 minutes, washed and mounted in Slowfade Diamond mounting oil (ThermoFisher).

For co-localisation analyses, cells were eventually incubated in culture with 25 nM MitoTracker (Molecular Probes) for 25 minutes before being fixed and processed for immunostaining. Alternatively, to label the cell surface, PCF were incubated with 10 μg/mL ConA-FITC for 1 hour at RT along with the secondary antibody solution during the immunostaining procedure.

Table 2.5. Antibodies used in immunofluorescence assays.

Antibody	Antigen	Type	Host	Dilution
α-MISP	*T. brucei* MISP (all)	Polyclonal purified	Rabbit	1:200
α-BARP	*T. brucei* BARP (all)	Polyclonal purified	Rabbit	1:200
α-CRD	*T. brucei* CRD	Polyclonal purified	Rabbit	1:50
α-HA	HA epitope	Polyclonal purified	Mouse	1:200
α-EP	*T. brucei* EP procyclin	Monoclonal purified	Mouse	1:800
α-PFR	*T. brucei* PFR	Polyclonal purified	Rabbit	1:100
α-rabbit-AF555	Rabbit IgG (H+L)	Polyclonal purified	Goat	1:500
α-mouse-AF488	Mouse IgG (H+L)	Polyclonal purified	Goat	1:500
α-mouse-AF555	Mouse IgG (H+L)	Polyclonal purified	Goat	1:500

2.5.2 Live immunostaining

Live cells were washed in ice-cold PBS and incubated in 20% FBS vPBS for 30 minutes on ice. The primary anti-MISP antibody was added (1:200 dilution) and incubated for 1 hour on ice. Cells were washed and incubated with the secondary goat anti-rabbit-Alexa Fluor-555 (1:500 dilution) for 1 hour on ice. Cells were then fixed in fresh 4% PFA for 15 minutes at RT, stained with DAPI 300 ng/mL for 10 minutes and mounted on a slide with Slowfade Diamond mounting oil for imaging.

2.5.3 Laser Scanning Confocal Microscopy (LSCM)

Specimens were imaged using the confocal laser scanning microscope Zeiss LSM-880 (located at the LSTM Cat 3 facility) with a Zeiss Apochromat 63X/1.4 objective and analysed with the Zen software (Zeiss). Imaging was carried out with lasers exciting at 405 nm, 488 nm, 514 nm, 561 and 633 nm according to the fluorochrome under excitation. The pinhole aperture was set by default at 1 unit, laser power and digital gain were adjusted to prevent saturation and scanning times were set for optimal results. For z-stack acquisition, the outer limits were set outside the specimen's surface and the number of stacks set to optimal. To measure mean intensity fluorescence in Zen, images were taken conserving identical laser power, pinhole aperture and digital gain settings. Cell length measurements were obtained using Fiji (ImageJ) from microscopy images.

2.5.3 Fluorescence-Activated Cell Sorting (FACS)

T. b. brucei PCF cells were washed in PBS and incubated with or without ConA-FITC 1 μg/mL in PBS for 1 hour on ice. Lectin control samples were incubated with pre-saturated ConA-FITC (pre-incubated with 200 mM α-methyl-D-mannoside and 200 mM α-methyl-D-glucoside at 27°C for 30 minutes). Tunicamycin-treated PCF controls had been previously cultured for 5 days in SDM-79 10% FBS supplemented with 1 μg/mL tunicamycin, a known *N*-glycosylation inhibitor (Takatsuki et al., 1971). After washing, cells were set at 10^6 cells/mL in PBS and acquired in a BD LSR II FACS system using the 488 nm laser and the FITC 530/30 detector. 20,000 cells were

acquired per sample at medium speed. Data were analysed using the FlowJo (FlowJo LLC) software.

2.6 Ion-exchange chromatography

For its activation, 8 g D-52 cellulose matrix (Whatman, GE) were incubated in 0.5M HCl for 30 minutes and washed in dH_2O until a pH=4 was reached. The matrix was then incubated in 0.5 M NaOH for 30 minutes and washed to reach a pH=8. The matrix was then washed in separation buffer (44 mM NaCl, 57 mM Na_2HPO_4, 3mM KH_2PO_4, pH=8) and pH was adjusted to 8 with orthophosphoric acid. Ion-exchange columns were prepared with ~0.5 mL matrix and equilibrated in separation buffer. *T. b. brucei cells* were washed with separation buffer supplemented with 55 mM D-glucose before being passed through the column.

2.7 Bioinformatics

Trypanosome sequences were obtained from TritrypDB (www.tritrypdb.org) (Aslett et al., 2010) or UniProt (www.uniprot.org) (Apweiler et al., 2004). *G. m. morsitans* sequences were obtained from VectorBase (www.vectorbase.org) (Lawson et al., 2007). Sequence analyses and multiple alignments were done using Clustal Omega (Sievers et al., 2011) and Geneious R9 (Biomatters Ltd). Alignment figures were made in Jalview (Waterhouse et al., 2009) and Geneious. Phylogenetic trees were generated using PyML 3.1 (Guindon and Gascuel, 2003) under default conditions with a set bootstrap of 500. Signal peptide, GPI-anchor signal, and transmembrane domain predictions were determined under default conditions using the servers SignalIP 4.1 (Petersen et al., 2011), PredGPI (Pierleoni et al., 2008) and TMHMM 2.0 (Krogh et al., 2001), respectively. *N*-glycosylation site predictions were run using the NetNGlyc 1.0 server (http://www.cbs.dtu.dk/services/NetNGlyc). Domain searches were done in InterPro (Apweiler et al., 2001). Protein structure modelling was performed using I-TASSER (when possible, providing a template crystal structure and sequence alignment) (Roy et al., 2010; Yang et al., 2015; Zhang, 2008) and IntFOLD (McGuffin et al., 2015). Available protein structures were obtained from Protein Data Bank (PDB, www.rcsb.org) (Bernstein et al., 1977). Structural homology searches were performed using DALI (Holm and Rosenstrom, 2010) and

conservation models using ConSurf (Ashkenazy et al., 2016; Ashkenazy et al., 2010). Protein structures were visualised and prepared for publication using PyMOL (Molecular Graphics System, Version 2.0 Schrödinger, LLC).

2.8 Statistics

Statistical analyses were applied whenever possible to data involving at least two technical or biological replicates. The analysis of variance (ANOVA) t-test was performed to analyse the significance between means of independent (unpaired) groups. Results with p-values <0.05 were considered significant. Typically, the means were plotted, and the standard deviations were shown as error bars using Prism (GraphPad).

CHAPTER 3

Metacyclic Invariant Surface Proteins (MISPs),
a novel family of trypanosome surface proteins

3.1 Introduction

All blood stages of *T. brucei* (i.e. the metacyclic and bloodstream trypomastigotes) are coated with a dense layer of VSGs which, in BSFs, are known to be responsible for antigenic variation (Horn, 2014) and clearance of VSG-bound host antibodies (Engstler et al., 2007). The VSG coat is also important because it masks other invariant surface proteins, hiding them from the host immune system. Thus, VSGs and antigenic variation are the main reason that has prevented the development of a vaccine against *T. brucei* and other African trypanosomes. Consequently, there is the need to identify, if any, new invariant proteins exposed over the VSG coat of either parasite stage which could lead to the development of a preventive (neutralising MCFs) or therapeutic vaccine (assisting the immune system to clear BSFs). As it is suggested in multiple studies searching for invariant surface proteins, it is unlikely that BSFs expose any conserved epitopes above the VSG coat as this would compromise the efficacy of antigenic variation (Schwede et al., 2011; Schwede et al., 2015). Metacyclics, however, may not need to fully protect invariant surface proteins since they are pre-adapted to survive in the host for a few days after transmission, and rapidly differentiate into BSFs once they get transmitted by the tsetse. They may not be under pressure to mask conserved surface proteins because this short period of time is not long enough for the host to effectively develop a specific immune response. In addition, MCF are transmitted to the skin of the host through the bite of the tsetse (Caljon et al., 2016). These are inoculated along with saliva, whose main function is to inhibit the host's haemostatic response to the cutaneous trauma, and it may also contain parasite factors important for transmission and early stages of infection in the host (Alves-Silva et al., 2010; Van Den Abbeele et al., 2010).

Altogether, studying the MCF surface and searching for parasite soluble factors in saliva seem to be reasonable approaches to follow in pursuit of the identification of new vaccine candidates against trypanosomiasis. This chapter describes a proteomics approach on *T. brucei*-infected tsetse saliva, that led to the identification of a novel family of *T. brucei* proteins I hereby characterised and named Metacyclic Invariant Surface Proteins (MISP). MISP is a small family of highly conserved glycoproteins from *T. brucei* (including *T. b. gambiense* and *T. b. rhodesiense*), *T. evansi* and *T. congolense*. Due to their unique characteristics, MISP are proposed to be molecular markers for metacyclogenesis, to have potential uses in xenodiagnosis in the field, and to be potential vaccine candidates against several African trypanosomiases.

3.2 Results

3.2.1 Mass spectrometry of trypanosome-infected tsetse saliva

3.2.1.1 Obtaining naïve and *T. brucei*-infected tsetse saliva samples[1]

Groups of tsetse flies infected with *T. b. brucei* BSF strain TSW-196, and control flies (naïve) of the same age were dissected at 30 days post infection (dpi) and saliva was collected, in three biological replicates. The rate of SG infections is usually ~10-20% under these experimental conditions. SG infections were assessed by microscopy (attached EMFs and MCF can be seen moving through the SG). The collected SGs were pooled and centrifuged to collect the supernatant enriched with saliva (pellet contained SG tissue and trypanosomes). Infected and naïve saliva samples were lyophilised, submitted to in-solution trypsin digestion, and analyzed by nano-liquid chromatography tandem mass spectrometry (nLC-MS/MS) (Dundee University Fingerprints Proteomics Facility).

3.2.1.2 MS/MS spectra analyses

The resulting MS/MS were searched with Proteome Discoverer against combined databases of *T. brucei* and *G. morsitans* proteins, and the resultant dataset was analysed using Scaffold Q+ and perSPECtives. After removing all *Glossina* hits and having manually inspected and curated all trypanosome hits and spectra, the resulting proteins were BLAST searched using BLAST2GO and mapped according to gene ontology (GO) terms. The complete workflow of sample processing is summarised in **Figure 3.1.**

---

[1] All the sample collection and initial proteomics data was obtained by Dr. Samïrah Perally. In collaboration with Prof. Igor Almeida (UTEP, USA), I then searched for significant hits, including manual curation of selected mass spectra, as shown below.

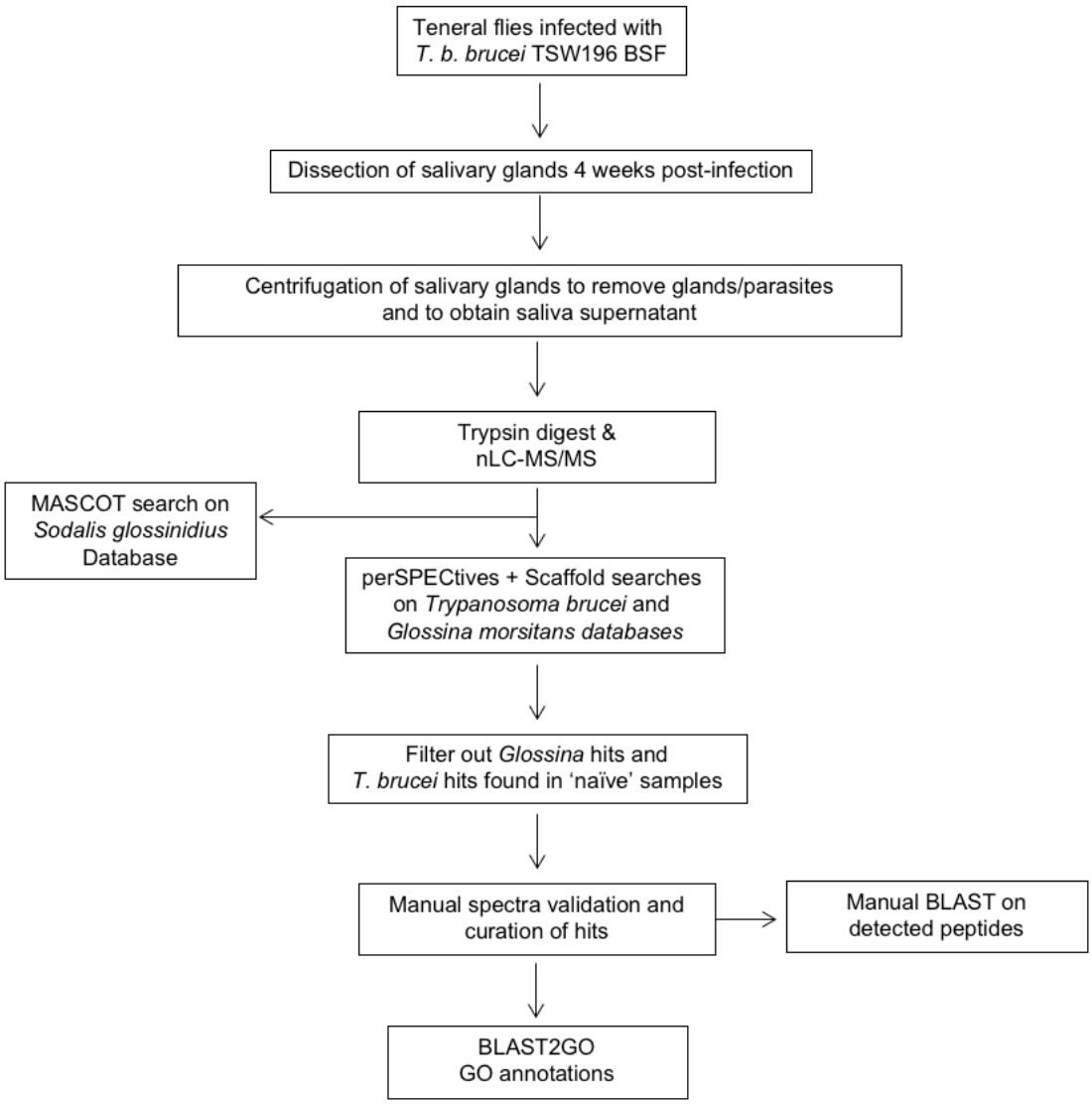


Figure 3.1. Workflow for the analysis of *T. brucei*-infected tsetse saliva. For each biological replicate, saliva samples from naïve flies were also collected and processed in parallel.

3.2.1.3 *T. brucei* proteins in infected tsetse saliva

We identified 45 unique peptides that derived from a total of 27 *T. b. brucei* proteins in infected saliva datasets, with high confidence identification (>95%). 17 of those proteins (62.9%) (**Figure 3.2, Table 3.1**) are intracellular mainly linked to functions such as protein folding (14.8%), cytoskeleton (11.1%) and metabolic processes (7.4%). The remaining 10 proteins (37.1%), were predicted to be GPI-anchored surface proteins from three major families: BARP, VSG and MISP (**Figure 3.2, Table 3.2**).

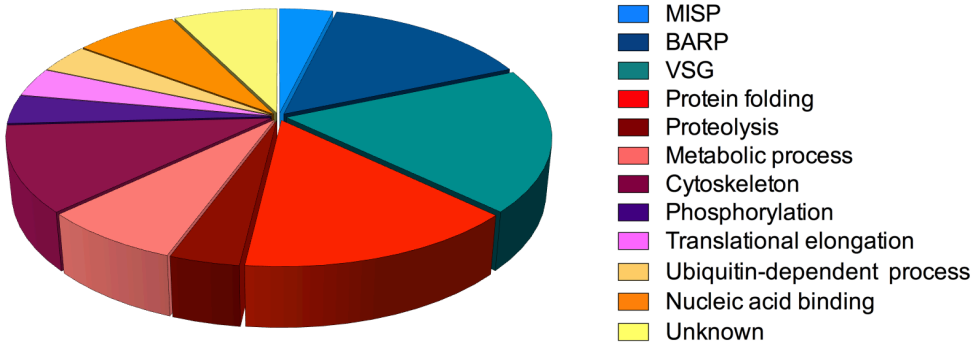


Figure 3.2. Pie chart of *T. brucei* proteins identified in tsetse saliva. Surface proteins (blue shades) and intracellular proteins (red to yellow shades) grouped based on functional GO terms.

Table 3.1. Intracellular *T. brucei* proteins identified by nLC-MS/MS in infected tsetse saliva. All proteins and peptides identified have >95% confidence and were manually curated. Accession code (Protein ID), protein description (Annotation), number of identified unique peptides (Peptides) and number of unique spectra (Spectra).

Protein ID [a]	Annotation	Peptides	Spectra
Tb927.1.2390	Beta tubulin	8	12
Tb927.11.11330	Heat shock protein 70	4	4
Tb927.10.5620*	Fructose bi-phosphate aldolase	2	2
Tb927.9.6210*	Arginine kinase	2	2
Tb927.10.4560	Elongation factor 2	2	2
Tb927.10.10280*	Microtubule-associated protein	2	2
Tb927.10.10460*	Histone 2B, putative	2	2
Tb927.11.3510	Peptidylpropyl isomerase	1	1
Tb927.3.4850	Enoyl-CoA hydratase, mitochondrial, putative	1	1
Tb927.11.2470	Metallo-peptidase, Clan MF, Family M17	1	1
Tb927.4.2740	p-25 alpha, putative	1	1
Tb927.10.10890*	Heat shock protein, putative	1	1
Tb927.2.2440*	Proteasome regulatory non-ATPase subunit 6	1	1
Tb927.10.6510*	Chaperonin HSP60, mitochondrial	1	1
Tb927.4.1340	Cleavage and polyadenylation specific factor subunit, putative	1	1
Tb927.8.6370	Cytoskeleton associated protein, putative	1	1
Tb927.3.5530	Tb-292 membrane associated protein	1	1

[(1)]: Accession code of coding gene in strain TREU927, TriTrypDB.
*: Representative ID of multiple homologs sharing the detected peptide.

Table 3.2. Summary of *T. brucei* GPI-anchored surface proteins detected in infected tsetse saliva by nLC-MS/MS. All proteins and peptides identified with >95% confidence and manually curated. Protein accession code (Protein ID), protein description (Annotation), number of identified unique peptides (Peptides), and number of spectra (Spectra).

Protein ID [a]	Annotation	Peptides	Spectra
Tb927.9.15520 [1]	Brucei Alanine-Rich Protein (BARP)	1	3
Tb927.9.15530	Brucei Alanine-Rich Protein (BARP)	4	7
Tb927.9.15570	Brucei Alanine-Rich Protein (BARP)	1	2
Tb927.9.15630	Brucei Alanine-Rich Protein (BARP)	1	1
VSM2_TRYBB	Variant Surface Glycoprotein (VSG) 221	3	4
O76421_TRYBR	Metacyclic Variant Antigen Type (mVAT) 4	1	1
A0A1J0RB71_9TRYP	Variant Surface Glycoprotein (VSG) 1125.4959	1	1
A0A1J0R978_9TRYP	Variant Surface Glycoprotein (VSG) 1125.3088	1	1
A0A1J0RAQ3_9TRYP [2]	Variant Surface Glycoprotein (VSG) 1125.408	1	1
Tb927.7.360 [3]	Metacyclic Invariant Surface Protein (MISP)	1	2

[a] Accession code for *T. b. brucei* in TriTrypDB (MISP and BARP) or UniProt (VSG).
[1] Representative ID. Peptide conserved in Tb927.9.15520, Tb927.9.15530, Tb927.9.15550, Tb927.9.15590, Tb927.9.15600, Tb927.9.15620 and Tb927.9.15660.
[2] Representative ID. Peptide also conserved in *T. b. brucei* VSGs 646, 769, 3613, 1125.408, 1125.474, 1125.1142, 1125.4207 and 1125.4707.
[3] Representative ID. Peptide conserved across all 5 *T. brucei* MISP isoforms.

3.2.1.4 *T. brucei* surface proteins: BARPs

One family of surface proteins identified is BARP, which are known to be expressed by EMFs in tsetse SG (Urwyler et al., 2007). According to the genome of *T. b. brucei* strain TREU927 (TriTrypDB), there are 14 *barp* homolog genes arranged in tandem (Berriman et al., 2005) (**Figure 3.3**). Out of 4 unique peptides detected, we identified 8 putative BARP isoforms (**Table 3.2**). For example, the peptide VQAEEAVELAESK (**Figure 3.4**) is conserved across 7 of the 14 BARP proteins (**Figure 3.3**). The peptide GLSQTVTSAVETSAAASSK is conserved in two isoforms (Tb927.9.15530 and Tb927.9.15630) and peptides QNTGWEVLTK and SAAEVVTAESISAALNDLGK are both unique for the isoform Tb927.9.15530.

3.2.1.5 *T. brucei* surface proteins: VSGs

We detected 7 unique VSG peptides (**Table 3.2**) that originated from the mammalian-infective MCFs and allowed the putative identification of VSG-221 (QADAANNFHDNDAECR, DSEASEIQTELK, LQLSPIQPK), mVAT4 (EAADVPCVNNGGGLNK), VSG 1125.4959 (LQQQLALYR), VSG 1125.3088 (STEELCNAK) and the peptide KTDAADKTEEK, which is found in 8 different VSGs (e.g. VSG 1125.408) (**Figure 3.5**). A multiple protein alignment revealed that these VSGs only share 10.5% identity. Since 6 out of 7 peptides did not correspond to metacyclic VSGs but to BSF VSGs, an extensive search over published similar works was done for comparison (**Table 3.3**).

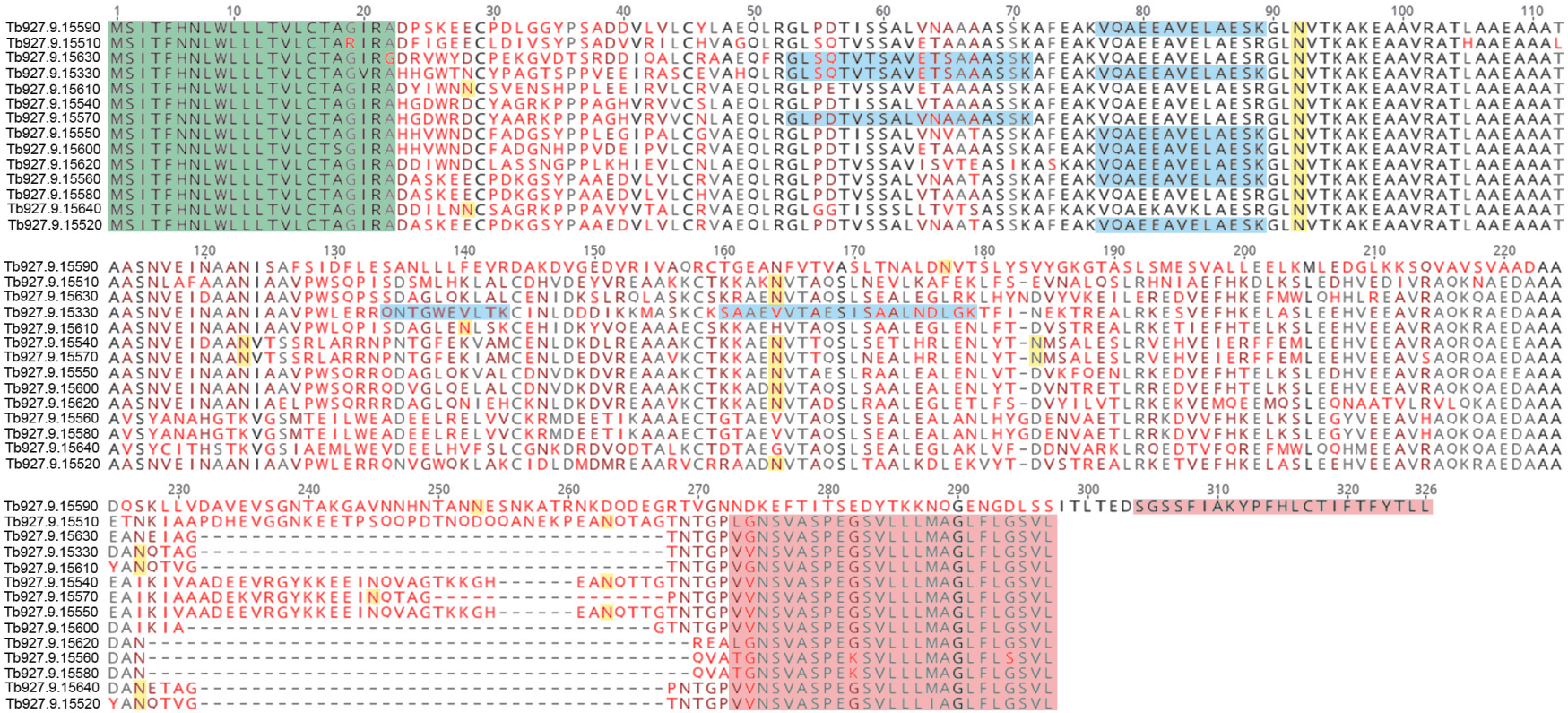


Figure 3.3. Multiple protein alignr

bright red (highest divergence). I

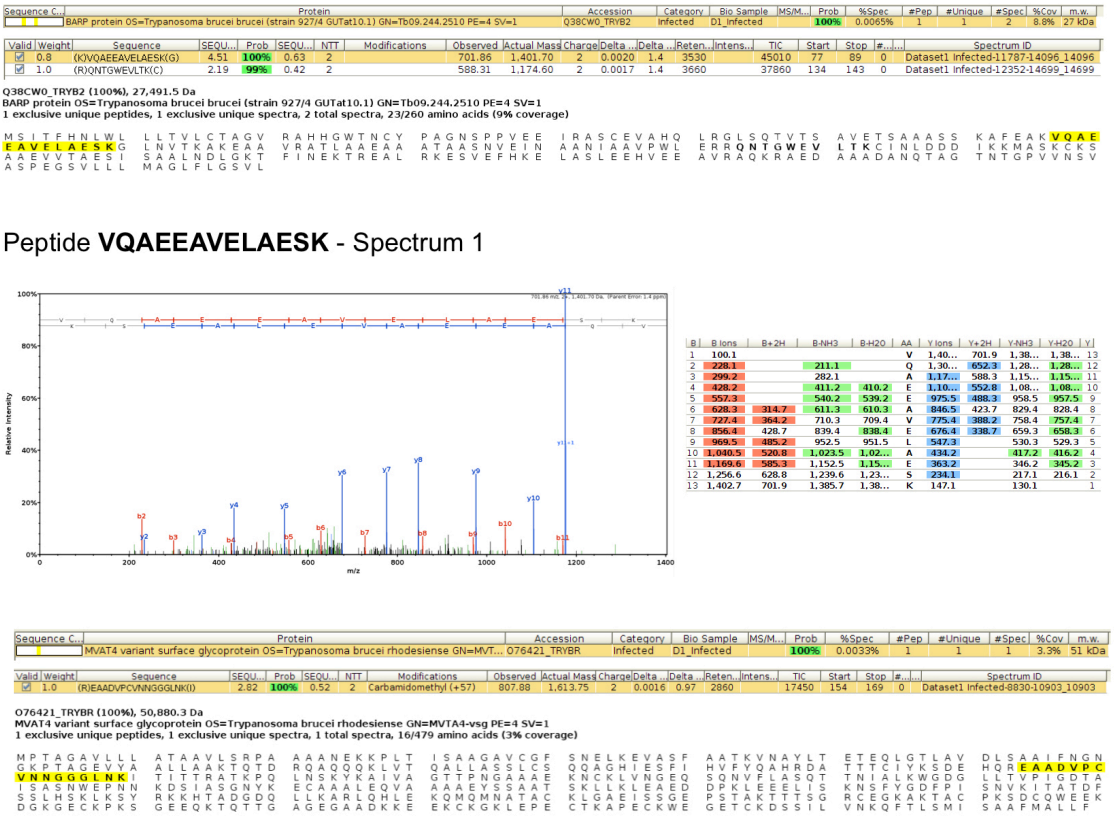


Figure 3.4. Example of mass spectrum of a *T. brucei* BARP peptide (A) and VSG mVAT4 (B) detected by nLC-MS/MS in tsetse infected saliva. Scaffold Q+ captures of identified protein, peptide coverage, biological sample source, spectra and ions table.

Although no study specifically reported finding BSF VSGs originating from MCF, it was found in the raw data that many of them identified BSF VSGs in either RNA or protein forms (Christiano et al., 2017; Kariithi et al., 2016; Savage et al., 2016; Telleria et al., 2014) (see Discussion).

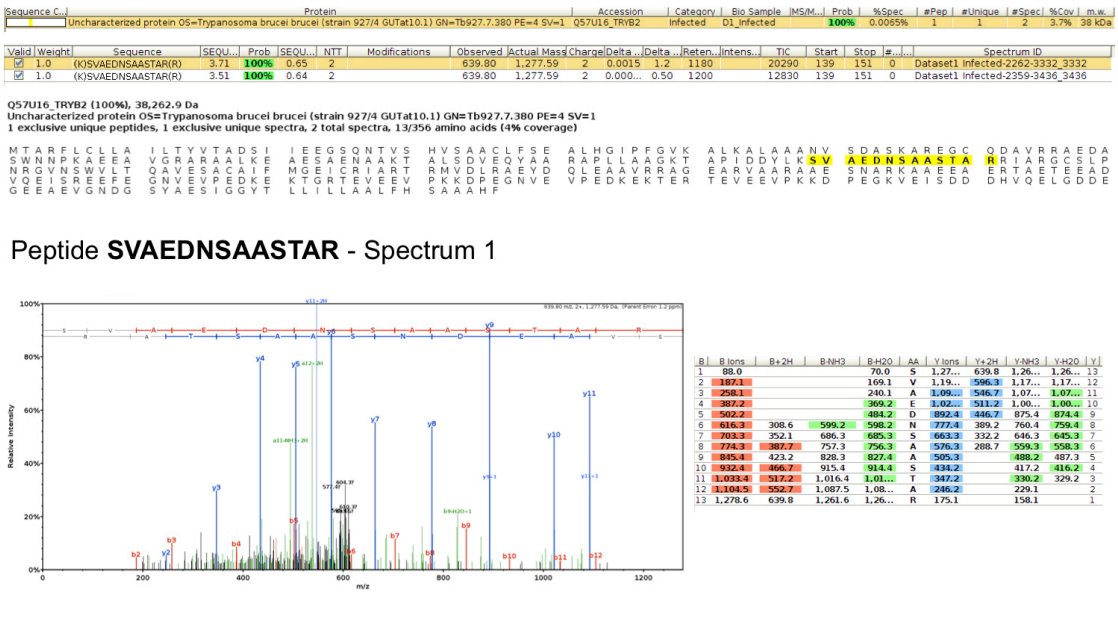
Table 3.3. Summary of *T. brucei* VSG species found in either cultured or fly-derived metacyclics (Literature).

Source [a,b]	RNA (RNA-seq) [c]	Protein (MS/MS) [c]	Reference
Tbb-infected tsetse salivary glands (RUMP 503)	Tb927.5.3990 (Q57Z50_TRYB2)	-	Telleria et al. 2014
Tbb-infected tsetse salivary glands (RUMP 503)	**VSG ILTat 1.22 (VSI2_TRYBB)** **VSG ILTat 1.61 (O97352_9TRYP)** **VSG ILTat 1.63 (Q8MPG1_9TRYP)** **VSG ILTat 1.64 (Q8MPG0_9TRYP)**	-	Savage et al. 2016
Saliva from *Tbb*-infected tsetse (EATRO 1125 AnTaR 1)	-	**mVAT5 (Q26842_9TRYP)** VSG 1228 (A0A1J0R6Q7_9TRYP) / VSG 1255 (A0A1J0R6U9_9TRYP) VSG 725 (M4SU87_9TRYP)	Kariithi et al. 2016
Cultured *T. brucei* MCF (Lister 427 29:13 *rbp6* overexpressor)	**VSG 397 (A0A1J0R4A4_9TRYP)** **VSG 531 (M4SYA9_9TRYP)** **VSG 639 (M4TDP9_9TRYP)** **VSG 653 (M4SYN2_9TRYP)** **VSG 1954 (M4T0T6_9TRYP)** Tb927.5.4690 / Tb927.9.1050 / Tb927.11.18330 [d] Tb927.1.5300 [e]	**VSG 397 (A0A1J0R4A4_9TRYP)** **VSG 531 (M4SYA9_9TRYP)** **VSG 639 (M4TDP9_9TRYP)** **VSG 653 (M4SYN2_9TRYP)** **VSG 1954 (M4T0T6_9TRYP)** Tb927.5.291b (B2ZWB3_TRYB2) Tb927.9.7380 / Tb927.1.5060 [f]	Christiano et al. 2017
Saliva from *Tbb*-infected tsetse (TSW-196)	-	**mVAT4 (O76421_TRYBR)** VSG 221(VSM2_TRYBB) VSG 4959 (A0A1J0RB71_9TRYP) VSG 3088 (M4SWM0_9TRYP) VSG 408/646/769/3613/ 474/1142/4207/4707 (M4SWZ7_9TRYP)	This thesis

Metacyclic VSGs in bold font
[a] *In vivo* samples isolated from infected *G. m. morsitans* flies
[b] *T. brucei* strain in parentheses
[c] VSG species; TritrypDB gene accession code; UniProt protein accession code
[d] Genes with higher differential expression (MCF compared to PCF) than canonical mVSGs (>1100 fold)
[e] There are 11 more VSG genes with differential expression MCF/PCF within the canonical mVSG range (250 to 730-fold)
[f] Also upregulated in RNA-seq data

3.2.1.6 *T. brucei* surface proteins: MISP

Lastly, we identified a unique peptide (SVAEDNSAASTAR) with 100% confidence (supported by two separate spectra; **Figure 3.5**), which belongs to a family of hypothetical GPI-anchored surface proteins (**Table 3.2**) that I further named Metacyclic Invariant Surface Proteins (MISP). The peptide SVAEDNSAASTAR is fully conserved across all MISP isoforms (see below; **Figure 3.9**).



Peptide **SVAEDNSAASTAR** - Spectrum 1

Figure 3.5. Example of mass spectrum of a *T. brucei* MISP peptide detected by nLC-MS/MS in tsetse infected saliva. Scaffold Q+ captures of identified protein, peptide coverage, biological sample source, spectra and ions table.

3.2.1.7 *Sodalis glossinidius* proteins in tsetse saliva

In addition to proteins from *T. brucei* and *G. morsitans*, it was also detected in both naïve and trypanosome-infected saliva a few peptides from the symbiont *Sodalis glossinidius* (Cheng and Aksoy, 1999) (not shown). The presence of *Sodalis* proteins was corroborated by western blotting using 1H1 monoclonal antibody, which specifically recognises *Sodalis* Hsp60 (Haines, 2002) (**Figure 3.6**). Taking together,

T. brucei-infected tsetse saliva contains a complex mixture of soluble factors from trypanosomes and *Sodalis* bacteria, in addition to a diminished proportion of *Glossina* salivary proteins compared to saliva from naïve flies (Van Den Abbeele et al., 2010).

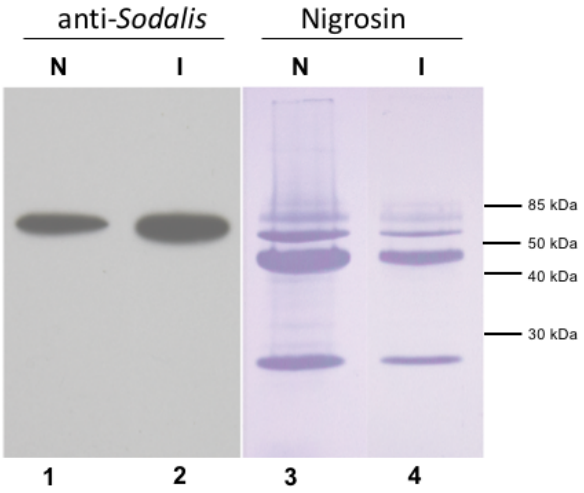


Figure 3.6. Immunoblotting of *G. m. morsitans* naïve and *T. b. brucei*-infected saliva probing with the mouse *anti-Sodalis* mAb 1H1. (1) Naïve (N) saliva and (2) infected (I) saliva probed with the antibody; film exposure: 10 seconds. (3) same naïve saliva and (4) infected saliva in PVDF membrane stained with nigrosin. Performed by Dr. Lee Haines (LSTM).

3.2.2 *In silico* analyses on MISP

3.2.2.1 Phylogenetic analyses

MISP were previously included in the cell-surface phylome (Jackson et al., 2013), within the large family Fam50 of trypanosome surface proteins which includes proteins such as the BARPs (Urwyler et al., 2007), and the *T. congolense* Glutamic Acid-Rich Protein (GARP) (Bayne et al., 1993; Beecroft et al., 1993) and Congolense Epimastigote Surface Protein (CESP) (Sakurai et al., 2008). The *T. b. brucei* genome contains five *misp* homolog genes (*Tb927.7.360, Tb927.7.380, Tb927.7.400, Tb927.7.420, Tb927.7.440* in strain TREU927, TriTrypDB) (Berriman et al., 2005), which encode for proteins that are unique in sequence and have no predictable

function. All five *T. b. brucei* homologs are found in an 'array-like' disposition in chromosome 7 (**Figure 3.7**), although the array contains unrelated genes between each pair of *misp* genes (i.e. *Tb927.7.370, Tb927.7.390, Tb927.7.410* and *Tb927.7.430*; all predicted to be homologs encoding the Golgi complex component 7). The same homologs are found in the human pathogenic subspecies *T. b. rhodesiense* and *T. b. gambiense* (*Tbg972.7.270*, *Tbg972.7.290*). Furthermore, slightly divergent versions of *misp* genes can also be found in the animal trypanosome species *T. evansi* (*TevSTIB805.7.300, TevSTIB805.7.320* and *TevSTIB805.7.380*) and *T. congolense* (*TcIL3000.0.02370*). A phylogenetic analysis using PhyML 3.1 (Guindon and Gascuel, 2003) on all MISP proteins determined that the family is divided in two sub-families, MISP-A and MISP-B, that originate from a common ancestor (**Figure 3.8-A**). The MISP-A sub-family, characterised by containing the N-terminal sequences Asp^{18}-Ser-Ile-Ile-Glu^{22} and Ala^{127}-Gly-Lys-Thr^{130}, and 1-3 repeats at the C-terminal (see below; **Figure 3.8-B**), is comprised by *T. b. brucei* Tb927.7.360, Tb927.7.380, Tb927.7.440, *T b. gambiense* Tbg972.7.290 and *T. evansi* TevSTIB805.7.300, TevSTIB805.7.320 and TevSTIB805.7.380 isoforms. MISP-B proteins, on the other hand, contain a conserved Gly^{18}-Pro-Lys-Gly-Gly^{22} and Asn^{127}-Asn-Asn-Val^{130} at the N-terminus and no C-terminal repeats, and is represented by Tb927.7.400 and Tb927.7.420 (*T. b. brucei*), and *T. b. gambiense* Tbg972.7.270. Other N-terminus residues (e.g. Ser^{25}, Ans^{27}, Phe^{38}, Arg^{121}, and Pro^{123}-Leu^{124} (for MISP-A), and Phe^{25}, Asp^{27}, Met^{38}, Asn^{121}, and Ser^{123}-Ala^{124} (for MISP-B)) are characteristic in each protein sub-family (**see Figure 3.8-B**). As for *T. congolense* TcIL3000.0.02370, it cannot be included within either group as it does not share the common ancestor and it only conserves a 34.2% of identity with other MISP isoforms. A comparison of the five *T. b. brucei misp* homologs yielded an overall sequence identity of 82.9% and 80.5%, at the nucleotide and amino acid level (**Figure 3.8-B**), respectively. In addition, the amino acid sequence alignment showed that the N-termini domains (Met^{1}–Glu^{237}) are more conserved (88.7%) than the C-termini (Glu–Lys^{238}) where the main divergence is found (49.0%). However, divergence is not random, but based on repetitions of negatively charged and acidic sequences of 26 residues (EVEEVPKKDPEGNVEVPEDKEKTERT or DVQEISREEFEGNVEVPEDK EKTERT). Alignment of all MISP proteins across species (excluding TcIL3000.0.02370) shows an overall 79.1% identity conservation (**Figure 3.8-B).**

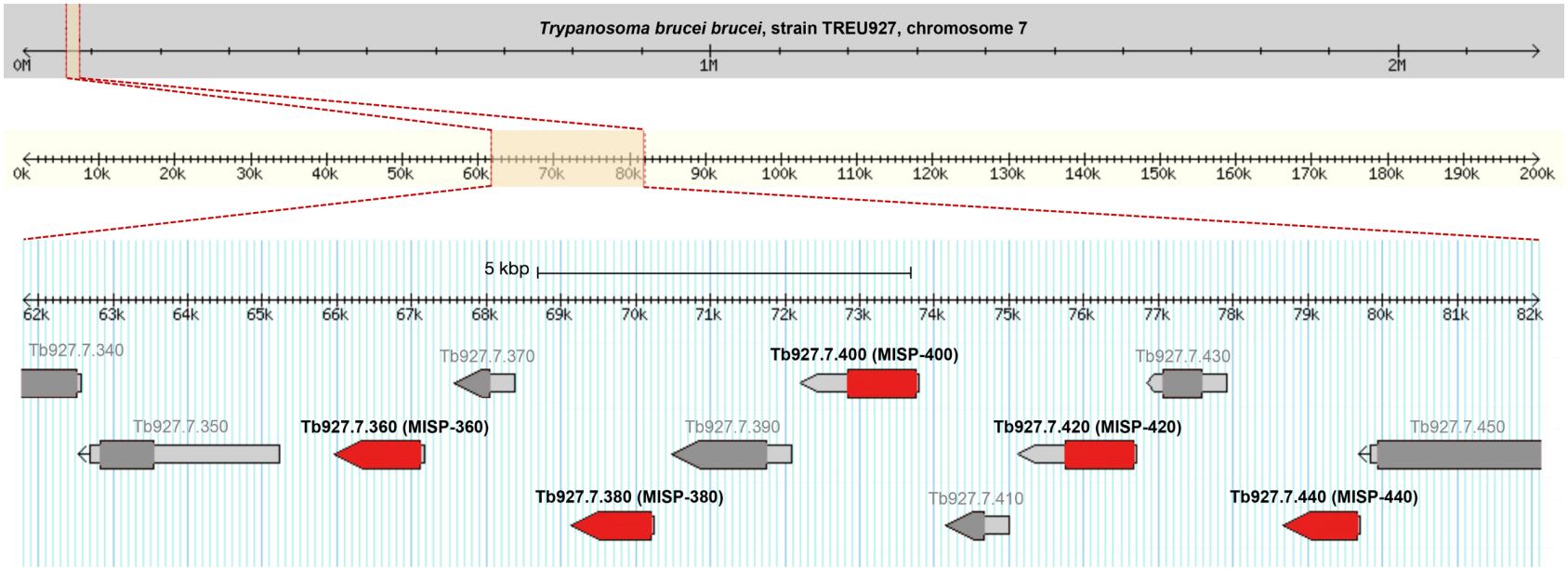


Figure 3.7. Schematic of the gen

TriTrypDB Genome Browser, from

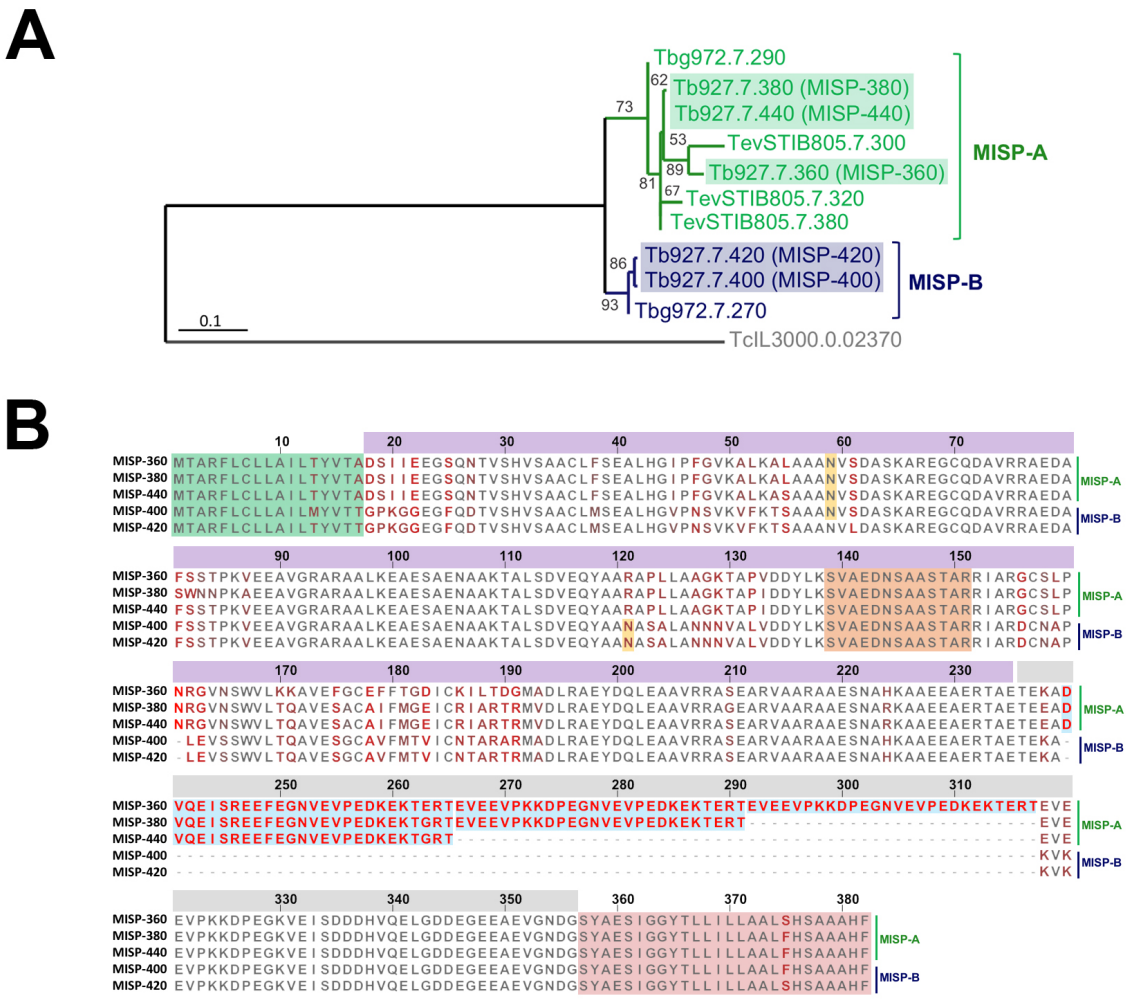


Figure 3.8. Sequence analysis of MISP protein family. (A) Unrooted phylogenetic tree estimated from a multiple alignment of all MISP proteins found across trypanosome species, pointing out the two sub-families MISP-A (green) and MISP-B (blue), and the *T. congolense* MISP (grey). The *T. b. brucei* MISPs are highlighted. Tree created using PhyML 3.1 and the WAG substitution model of maximum likelihood; bootstrap of 500; numbers show branch support values (%); scale bar = 0.1 substitutions/site. **(B)** Multiple protein alignment of the *T. b. brucei* MISP isoforms (MISP360, MISP380, MISP440, MISP400 and MISP420). Residues coloured from grey (full conservation) to bright red (highest divergence). Protein domains highlighted: signal peptide (green), putative *N*-glycosylation sites (yellow asparagine residues), C-terminus motifs of 26 residues (blue) and GPI-anchor signal peptide. The conserved peptide 'SVAEDNSAASTAR' identified by nLC-MS/MS is highlighted in orange. The top bar numbers indicate residue position starting from Met1. Top bar colours indicate the main protein domains N- (purple) and C- termini (grey).

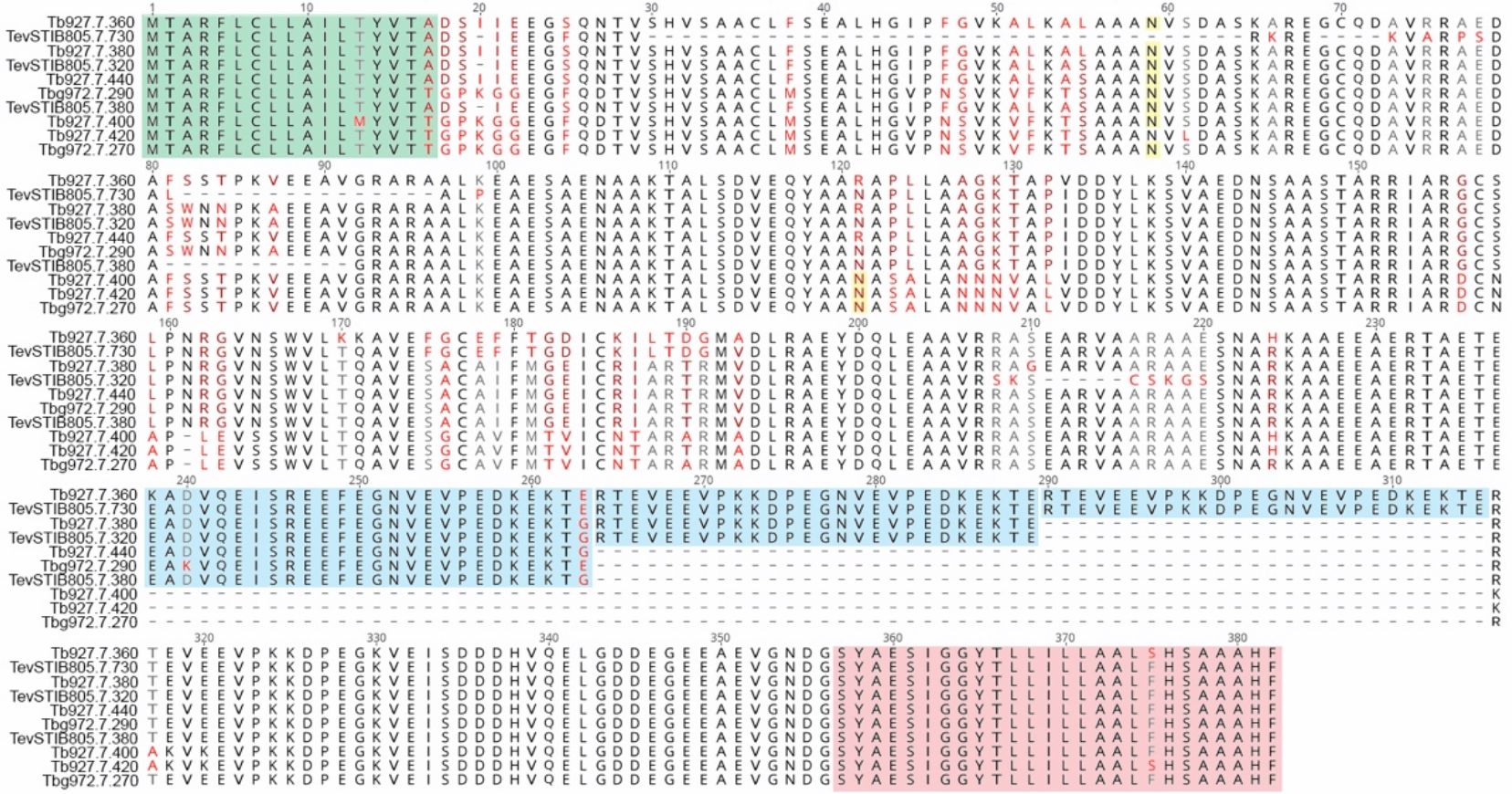


Figure 3.9 Protein multiple alignment

conserved) to bright red (most dive

3.2.2.2 Predictions on signal peptide, GPI-anchor and *N*-glycosylation

The presence of signal peptides in all *T. b. brucei* MISP isoforms (Met[1]–Ala/Thr[17]) was confirmed using the SignalP 4.1 server (Petersen et al., 2011), although, MISP-B Tb927.7.400 may also contain an alternative cleavage site at Gly[24] (**Table 3.4**). MISP proteins from other trypanosome species have equivalent signal peptide predictions (**Table 3.4**). According to the TMHMM 2.0 server, none of the MISP sub-families appear to contain a transmembrane domain (**Table 3.4**). However, all *T. brucei* isoforms have a strong prediction (99.9% specificity) for GPI-anchoring (Pierleoni et al., 2008), with a putative ω site at Ser within the C-terminal sequences SYAESIGGYTLLILLAALSHSAAAHF or SYAESIGGYTLLILLAALFHSAAAHF (**Figure 3.9-A, Table 3.4**). Furthermore, the NetNGlyc 1.0 server predicted that all *T. b. brucei* and *T. b. gambiense* MISP isoforms contain one *N*-glycosylation sequon (Asn[59]-Val-Ser; except for Tb927.7.420), and members of the MISP-B sub-family may also contain a putative secondary *N*-glycosylation sequon (Asn[121]-Ala-Ser). In addition, two *T. evansi* MISPs (TevSTIB805.7.320 and TevSTIB805.7.380) contain one *N*-glycosylation site (Asn[58]-Val-Ser) and, interestingly, the *T. congolense* MISP is potentially *N*-glycosylated at 6 different sites (Asn[126]-Ala-Thr, Asn[163]-Asp-Thr, Asn[166]-Ser-Ser, Asn[195]-Ile-Thr, Asn[198]-Ala-Thr, Asn[245]-Thr-Ser) (**Table 3.4**).

Table 3.4. I*n silico* predictions based on MISP protein sequences. Accession code (Protein ID), trypanosome species having the coding gene (Species), MISP sub-family (Sub-family), predicted signal peptide (SP), predicted GPI anchor peptide (GPI), number of predicted *N*-glycosylation sites (*N*-glycan) and number of C-terminus 26-residues motives (C- mot.)

Protein ID [a]	Species	Sub-family	SP [b]	GPI [c]	*N*-gly	CTR [d]
Tb927.7.360	*T. b. brucei / rhodesiense*	MISP-A	1-17	Yes	1	3
Tb927.7.380	*T. b. brucei / rhodesiense*	MISP-A	1-17	Yes	1	2
Tb927.7.400	*T. b. brucei / rhodesiense*	MISP-B	1-24	Yes	1-2	0
Tb927.7.420	*T. b. brucei / rhodesiense*	MISP-B	1-24	Yes	0	0
Tb927.7.440	*T. b. brucei / rhodesiense*	MISP-A	1-17	Yes	1	1
Tbg972.7.270	*T. b. gambiense*	MISP-B	1-17/24	Yes	1-2	0
Tbg972.7.290	*T. b. gambiense*	MISP-A	1-17/24	Yes	1	1
TevSTIB805.7.300	*T. evansi*	MISP-A	1-17	Yes	0	3
TevSTIB805.7.320	*T. evansi*	MISP-A	1-17	Yes	1	2
TevSTIB805.7.380	*T. evansi*	MISP-A	1-17	Yes	1	1
TcIL3000.0.02370	*T. congolense*	-	1-22	Yes	6-7	-

[a]: Accession code of coding genes in TriTrypDB.
[b]: Starting-ending residues. Peptide cleaved after ending residue.
[c]: All GPI anchor peptides are formed by the last 26 residues of the protein.
[d]: 26 residues motifs described in Figure 3.8-A.

3.2.2.3 Prediction on protease target sequences

Using the mature MISP360 and MISP400 protein sequences (representatives for MISP-A and MISP-B), potential sites of cleavage by proteases were predicted using the PROSPER server (Song et al., 2012) (**Figure 3.10**). This identifies target sites for 24 different protease families from the major types aspartic (A), cysteine (C), metallo- (M) and serine (S) proteases. The predictions show a similar profile of multiple potential cleavage sites in both *Tb*MISP360 and *Tb*MISP400. Considering the crystal structure of the MISP360 N-terminus (see section 3.2.6.1) it seems unlikely that these sites are accessible to proteases. However, sites at the C-terminus may be

actually susceptible to protease cleavage because of the lack of secondary structure (see section 3.2.6.3). The release of these surface proteins by C-termini cleavage, which maintains the protein attached to the plasmatic membrane, may explain the detection of MISP in the saliva proteomics, although the exact mechanism of MISPs released from the parasite surface remains to be investigated.

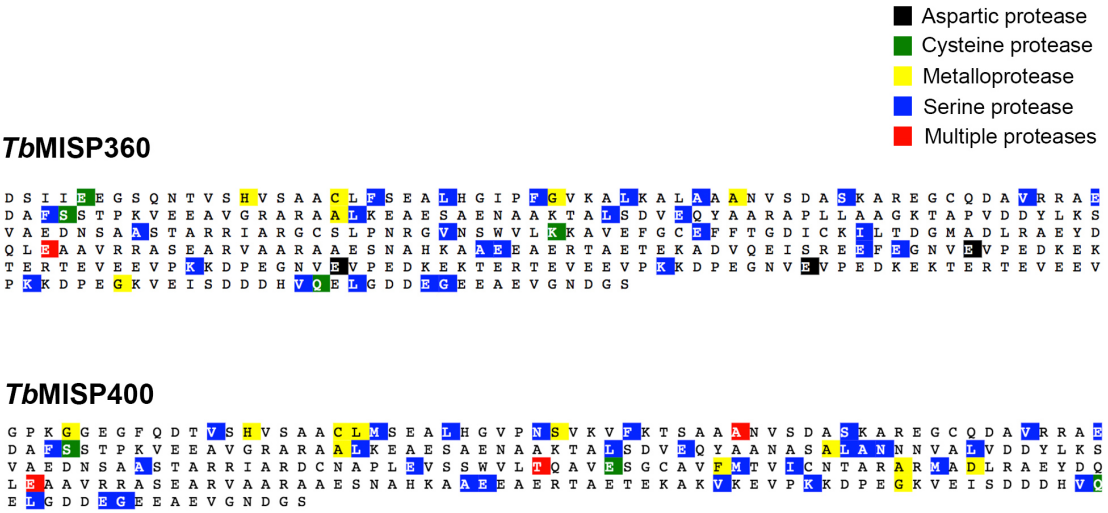


Figure 3.10. PROSPER protease cleavage sites predictions. MISP360 (top) and MISP400 sequences with predicted protease cleavage sites highlighted. Only one member of each family is represented. Key, the top right.

3.2.2.4 First insights into the 3D structures of MISP

To get a first idea into the possible protein 3D structure of MISP, the amino acid sequences of MISP360 (representing the MISP-A sub-family and having the longest C-termini) and MISP400 (representing MISP-B and the shortest C-termini), both lacking the signal and GPI-anchor peptides, were used in both I-TASSER (Roy et al., 2010; Yang et al., 2015; Zhang, 2008) and IntFOLD (McGuffin et al., 2015) servers for *in silico* protein structure modelling. Both servers found the MISP N-terminus to be structurally similar to the *T. congolense* GARP (PDB: 2y44) (Loveless et al., 2011). No structural hits were found for the C-termini, and MISP protein function

could not be predicted due to the lack of structural or sequence domain similarity. Final models of MISP360 and MISP400 were built in PyMOL 2.0 by assembling separate models of N- and C-termini (**Figure 3.11**). The generation and curation of these models was important to solve the crystal structure of MISP360 N-terminus by molecular replacement (**see section 3.2.6.1**) and for the design of recombinant versions of MISP360 for NMR studies (**see section 3.2.3.2**)

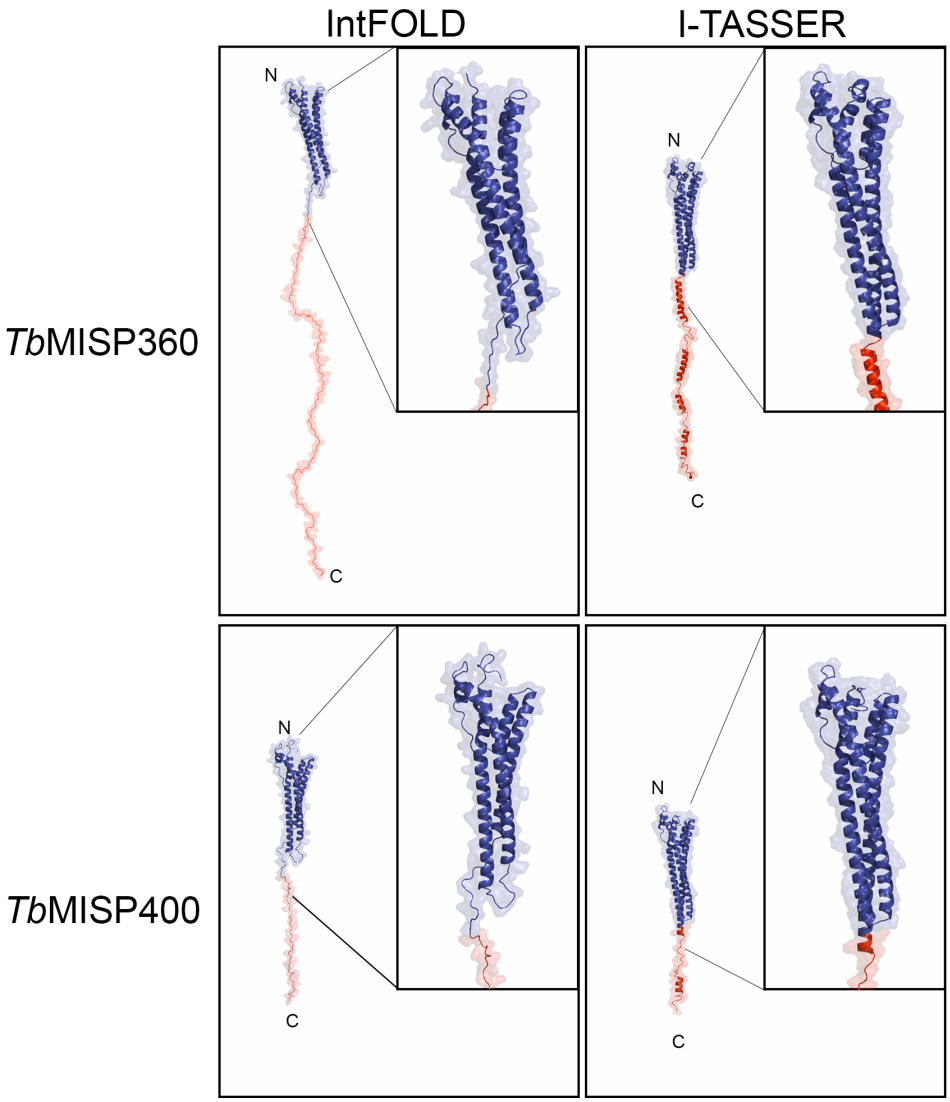


Figure 3.11. 3D structural models of *T. brucei* MISP360 and MISP400 proteins. Models generated with I-TASSER (left) and IntFOLD (right) servers. Cartoon representation with molecular surface of N-termini domains (blue) and C-termini domains (red). Insets show a close up of the N-termini domains.

3.2.3 Production of MISP recombinant proteins

3.2.3.1 Production of recombinant MISP for rabbit immunisation

In order to create an antibody against MISP (see section 3.2.4), a recombinant MISP (rMISP380) was produced for immunization[2]. The *misp380* gene was amplified by PCR from the *T. b. brucei* genome and cloned into the bacterial expression plasmid pET28. The complete coding sequence was depleted from the signal peptide and GPI-anchor signal peptide (both predicted to be absent in the mature protein) and fused to a poly-histidine tag (6xHis) at the C-terminus (**Figure 3.12-A**). The expression of the transgene is driven by a T7 promoter, and tightly regulated by a Lac operator sequence. The recombinant protein expression was induced with IPTG in *E. coli* and purified by nickel-agarose affinity (**Figure 3.12-B-C-D**). Several protein fractions of rMISP380 were collected and analysed by SDS-PAGE to select the elution fractions with the purest MISP content (**Figure 3.12-D**). All eluates were pooled together and further purified by size exclusion chromatography (SEC) using a FPLC system in a Superdex 200 column (**Figure 3.13**). Fractions were collected and analysed by Coomassie blue SDS-PAGE (**Figure 3.13-A**). The elution pattern showed that most of rMISP380 (predicted to be ~35 kDa in a monomeric state) comes out from the column with a molecular mass bigger than 650 kDa, suggesting that it may form non-specific aggregates that dissociate into monomers after processing for SDS-PAGE (e.g. fractions 5, 6, 7). A smaller peak at ~35 kDa suggests that rMISP380 also elutes as a monomer, but in a much-reduced proportion (fraction 19). Fractions 5 to 9 were pooled, yielding ~0.4 mg of purified rMISP380. This sample was used for the immunisation and antibody purification as described in Section 3.2.4.

---

[2] All attempts to express rMISP360 in bacteria, using the exact same procedure described for the expression of rMIPS380, were unsuccessful. However, due to the high degree of conservation among MISP isoforms, it is expected that antibodies anti-MISP380 cross-react with all MISP variants.

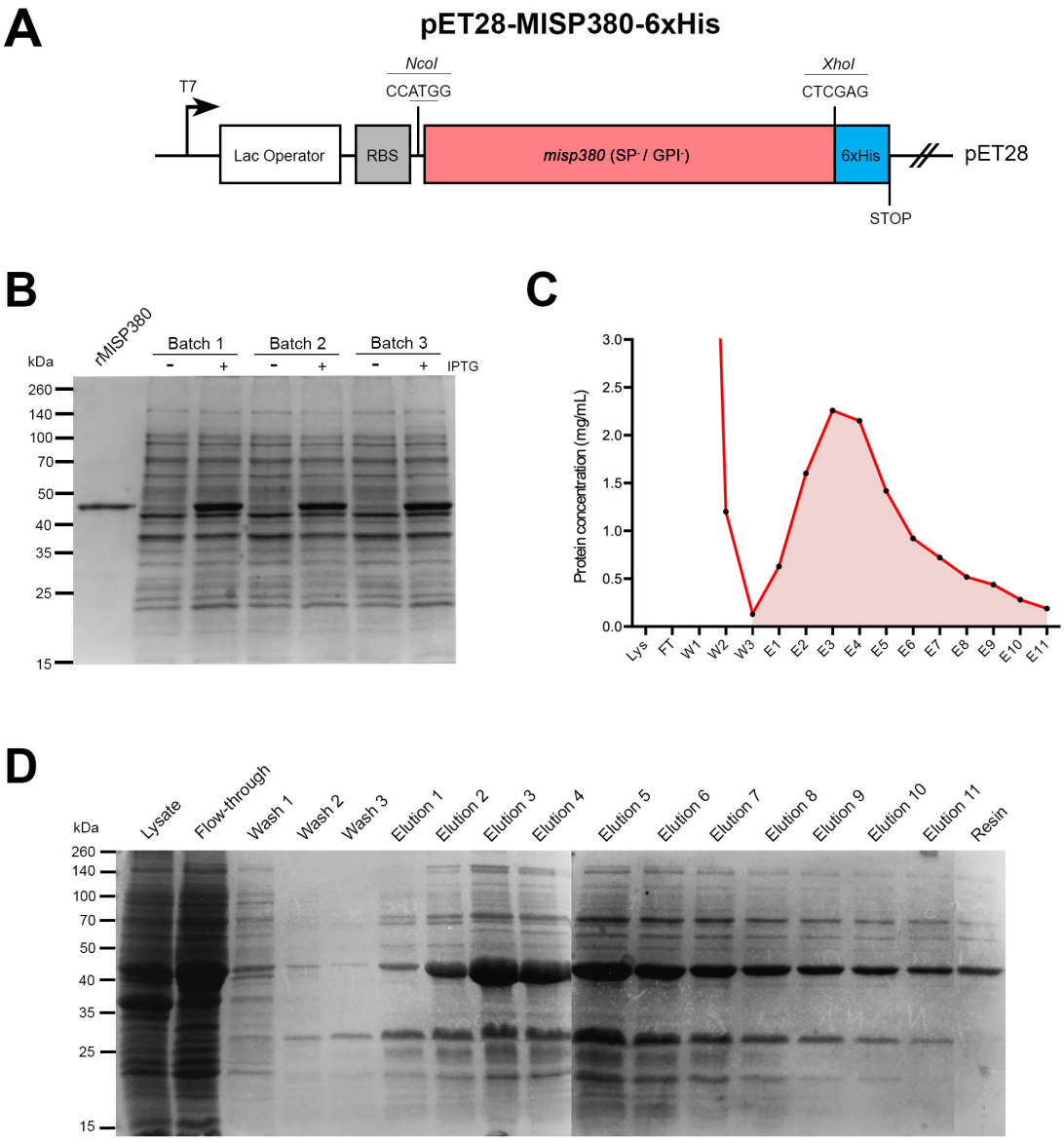


Figure 3.12. Expression and Ni-NTA purification of rMISP380. (A) DNA construct engineered for the expression of the protein. *misp380* gene (lacking signal and GPI peptides) cloned using *NcoI* (containing the start codon, underlined) and *XhoI* restriction targets. T7 promoter (T7), Lac operator and ribosome binding site found upstream the transgene. **(B)** SDS-PAGE of *E. coli* BL21(DE3) Rosetta 2 cell lysates of 3 different cultures induced (+) or not induced (-) with IPTG for the production of rMISP380, compared with the purified rMISP380. **(C)** Protein concentration (mg/mL) graph of the different fractions acquired during Ni-NTA purification: bacteria lysate (Lys), column flow-through (FT), washes 1-3 (W) and elutions 1-11 (E). **(D)** SDS-PAGE of Ni-NTA purification fractions.

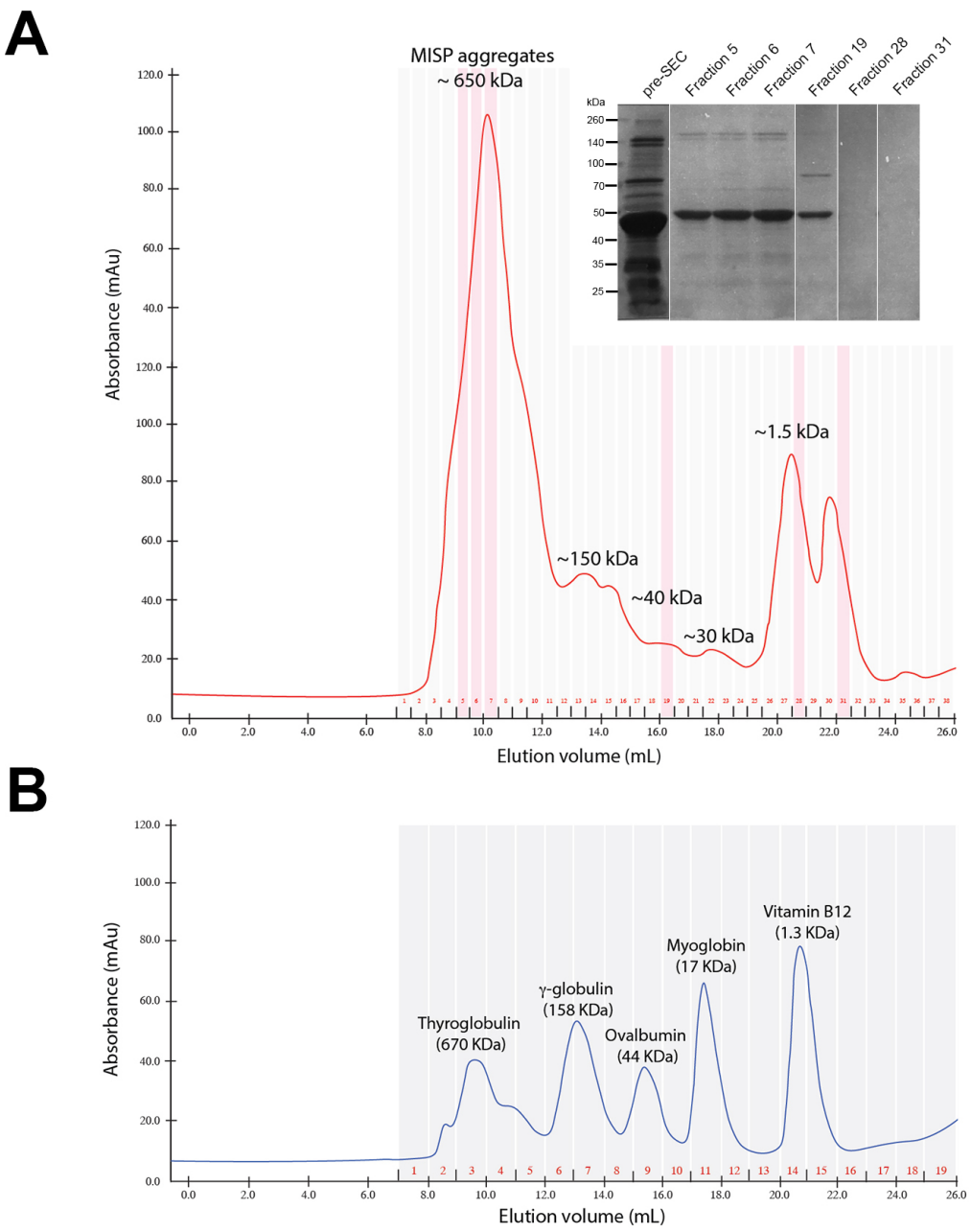


Figure 3.13. (A) SEC elution pattern of Ni-NTA-purified rMISP380 and fractions collected (red numbers). Estimated molecular weights shown above the major peaks. SDS-PAGE (inset) of the pre-SEC sample and selected fractions (highlighted in red). **(B)** SEC elution pattern of standards used in calibration and molecular size estimation.

3.2.3.2 Production of recombinant truncated variants of MISP360.

As shown in section 3.2.2.4, 3D modelling of MISP360 and MISP400 suggested that the C-terminus domain of MISP is likely to acquire an unstructured random coiled conformation, while the N-terminus domains seem to adopt a helical folding similar to the *T. congolense* GARP crystal structure (Loveless et al., 2011). The crystal structure of MISP360 N-terminus has been solved by X-ray crystallography (see section 3.2.6.1). However, the C-terminus could not be crystallised, probably due to this predicted disorganised nature, which makes this part of the protein very mobile. Accordingly, nuclear magnetic resonance (NMR) would be a more suitable method to solve the structure of the MISPs C-termini and/or the complete structures of MISP. Furthermore, it will also help to obtain more information about the motion dynamics of the different protein domains. Although due to time-constraints no NMR studies were conducted in this thesis, I produced a series MISP recombinant proteins that could be used for these purposes in the future.

The coding sequence of *misp360* was obtained by PCR from the *T. b. brucei* genome and cloned into a standard bacterial plasmid. From there, by using different primer pairs, the coding sequences for either the N- terminus, the C- terminus or the complete MISP360 proteins were cloned into the bacterial expression plasmid pCOLD II for production of recombinant protein under cold conditions to achieve an optimal protein folding (**Figure 3.14**). This is done by exploiting the *cspA* cold-shock protein promoter and UTRs. Coding sequences lacked the signal peptide and GPI-anchor signal peptide (removed in the mature protein) and were fused to a 6xHis tag at N-terminus.

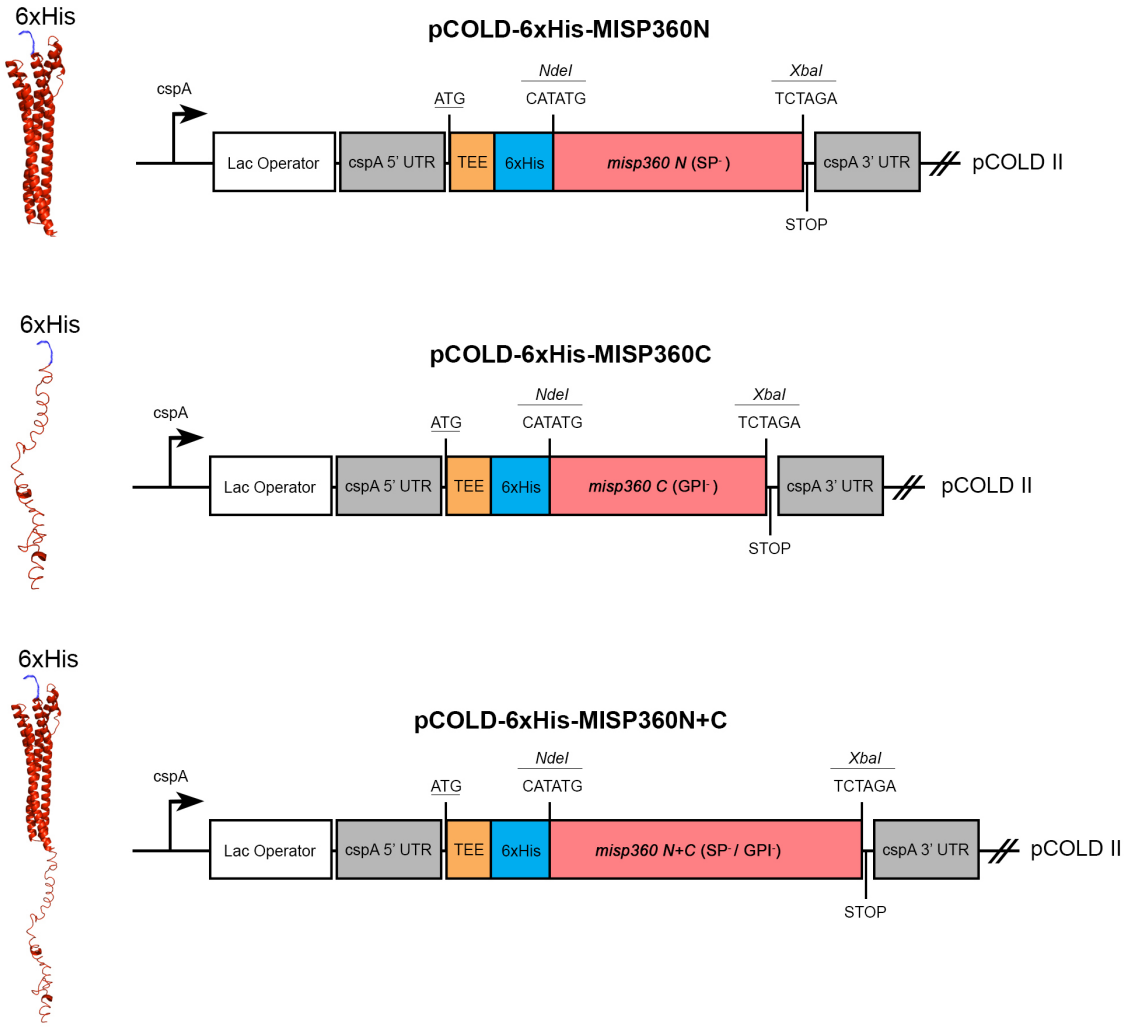


Figure 3.14. Engineered DNA constructs for the expression of MISP360 recombinant proteins in *E. coli*. (Left) 3D model of the resultant recombinant protein. (Right) DNA constructs coding for the recombinant protein. The 3 variations use the pCOLD II plasmid backbone for the inducible (IPTG) expression of the transgene under the cold-shock *cspA* promoter and UTRs. All proteins are 6xHis tagged at N-terminus and contain the translation enhancer element (TEE) upstream the tag.

The resultant proteins (^{N}MISP360, ^{C}MISP360 and $^{N+C}$MISP360) were purified by nickel-agarose affinity (**Figure 3.15**) and the eluted fractions that showed the highest purity by SDS-PAGE were pooled together and stored for further use (see reactivity with anti-MISP polyclonal in section 3.2.4). As seen in **Figure 3.15**, all 3 proteins were produced and had affinity for the purification column. ^{N}MISP360 production had the highest yield (measured by UV-Vis spectroscopy), followed by ^{NC}MISP360. Protein content of ^{C}MISP360 elutions could not be measured using UV-Vis spectroscopy since the protein lacks aromatic amino acids (Trp, Tyr and Phe) which absorb 280 nm light. ^{N}MISP360 is displayed as a major single band at ~25 kDa as its expected molecular weight. However, ^{C}MISP360 is shown in two bands with different molecular weights (~30 kDa and ~24 kDa) which do not correspond to the protein expected size (~15.6 kDa). Although proteins ran under denaturing conditions, this may indicate that the C-termini domains form several forms of dimeric aggregates. The complete ^{NC}MISP360 protein, similar to rMISP380 from section 3.2.3.1, is displayed as a major single band with a relative molecular size of ~50 kDa, much bigger than its expected weight (~38.6 kDa).

The presence of the unstructured C-terminus domain composed of low complexity regions seems to affect the yield of protein production and to alter the apparent molecular size on SDS-PAGE of both ^{C}MISP360 and ^{NC}MISP360 proteins. Interestingly, the two bands that appear to correspond to the MISP360 C-terminus are also shown in ^{NC}rMISP360 and rMISP380 purifications, but not in ^{N}MISP360 which does not contain a C-terminus domain. This suggests that the C-terminus may be very susceptible to cleavage or degradation, as predicted in Section 3.2.2.3.

3.2.4 Creation of a polyclonal antibody against MISP

A polyclonal antibody was raised against MISP as a valuable tool to study many aspects of these proteins. Since the *T. brucei* MISP family is highly conserved (shown above), it was hypothesised that raising a polyclonal antibody by immunisation with a complete MISP protein would lead to a highly cross-reactive antibody for all isoforms. This antibody was widely used to detect expression of MISP proteins by immunoblotting and for cell localisation studies by immunofluorescence.

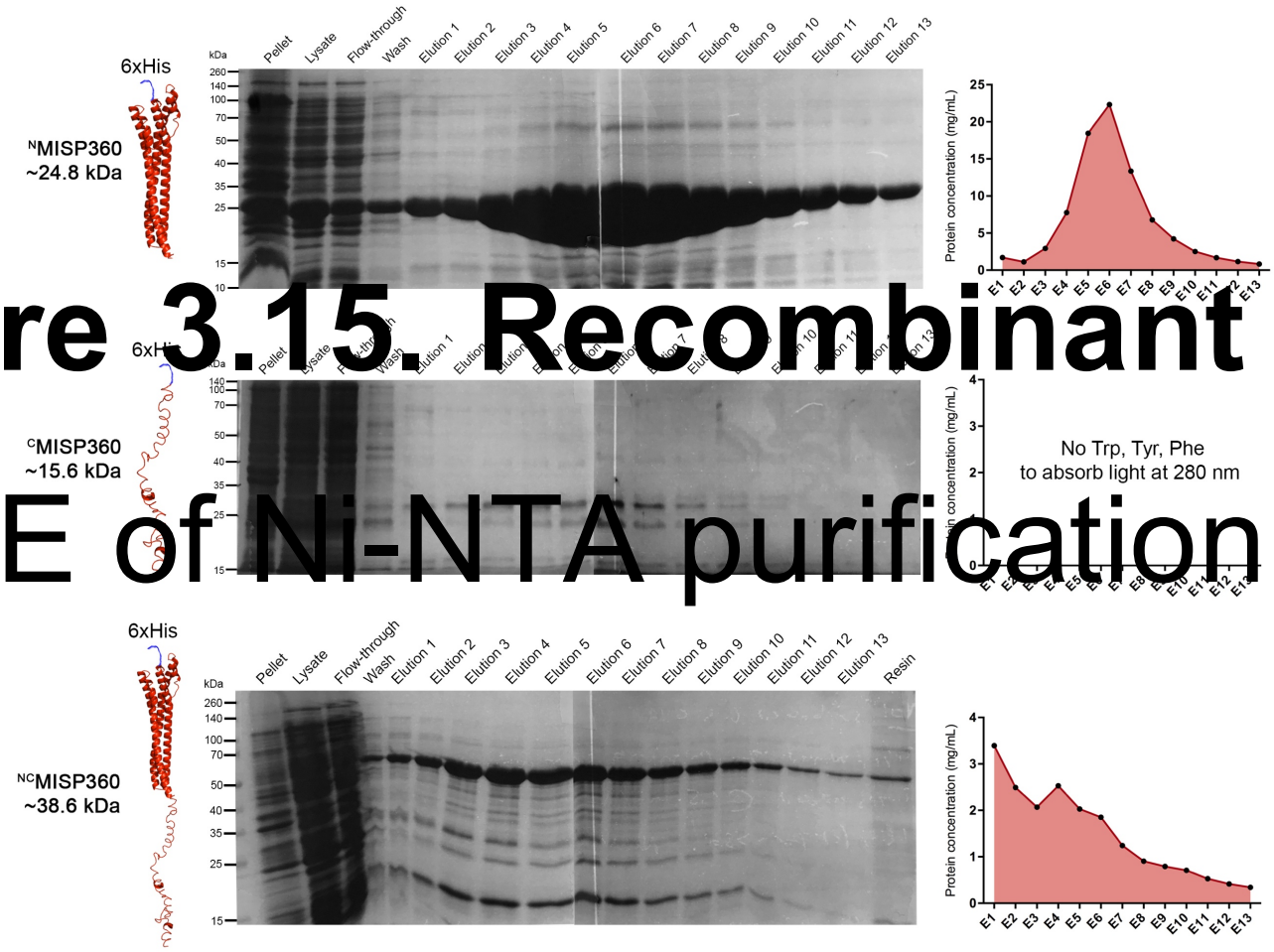


Figure 3.15. Recombinant MISP3

PAGE of Ni-NTA purification (right)

The immunisation and antibody purifications were carried out by the company Davids Biotechnologie GmbH (Germany) on a single rabbit using 0,3 mg of purified recombinant rMISP-380 (described in section 3.2.3.1), and anti-serum was obtained after bleeding the animal at 63 days post-immunisation (**Figure 3.16**). The anti-serum was then processed to purify IgGs using a protein A column. These IgGs were then affinity-purified using a column containing a matrix covalently bound to rMISP380. Lastly, anti-MISP IgGs were further passed through a column containing protein lysate from the same *E. coli* cells used to produce the recombinant protein, so IgGs with unspecific binding to bacterial proteins were excluded. A pre-immunisation rabbit serum was also acquired. All anti-sera and purified antibodies were tested for reactivity with the recombinant protein by immunoblotting (**Figure 3.17**). Reactivity was also tested after sample freezing, before aliquots were made and kept frozen at -20°C until further use.

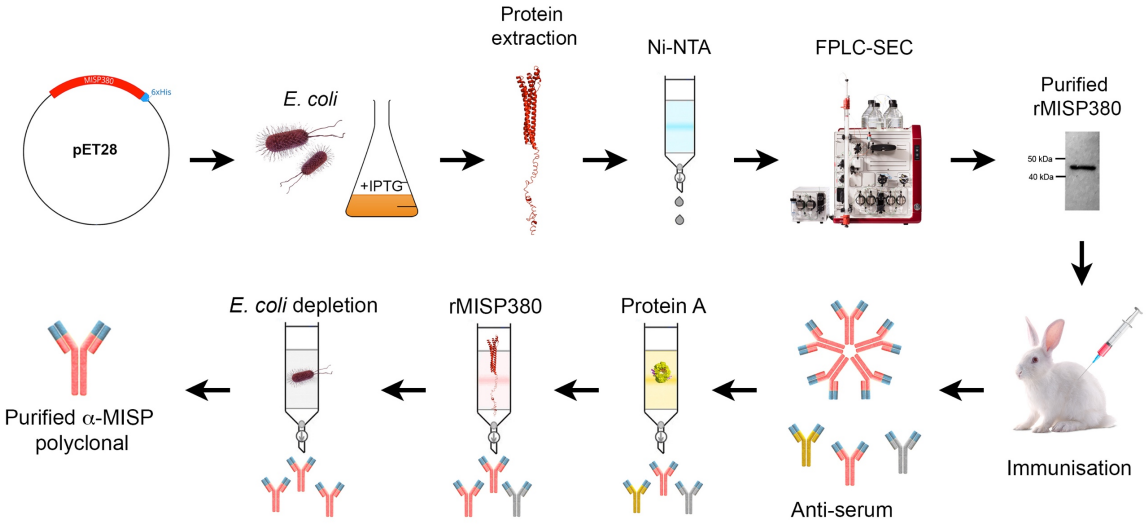


Figure 3.16. Schematic representation depicting the steps involved in the production of polyclonal antibodies α-MISP. Detailed procedure described in section 3.2.4

Several western blottings were performed to evaluate the reactivity and specificity of the polyclonal anti-MISP. The antibody was tested against the rMISP380 protein (used for the immunisation) as a positive control, bovine serum albumin as negative control, and the recombinant protein ^{N}MISP360 (see section 3.2.3.2) as test subject. As shown in **Figure 3.17**, the antibody positively reacted with several rMISP380 (from independent batches) and ^{N}MISP360, partly confirming cross-reactivity with at least one more isoform. Neither purified anti-MISP nor anti-serum reacted with a *T. brucei* PCF lysate, proving to have null unspecific binding to trypanosomal proteins and suggesting that MISP may not be expressed in PCFs. In addition, pre-immune serum did not react with any of the recombinant MISP variants, proving to be a good negative control in further assays.

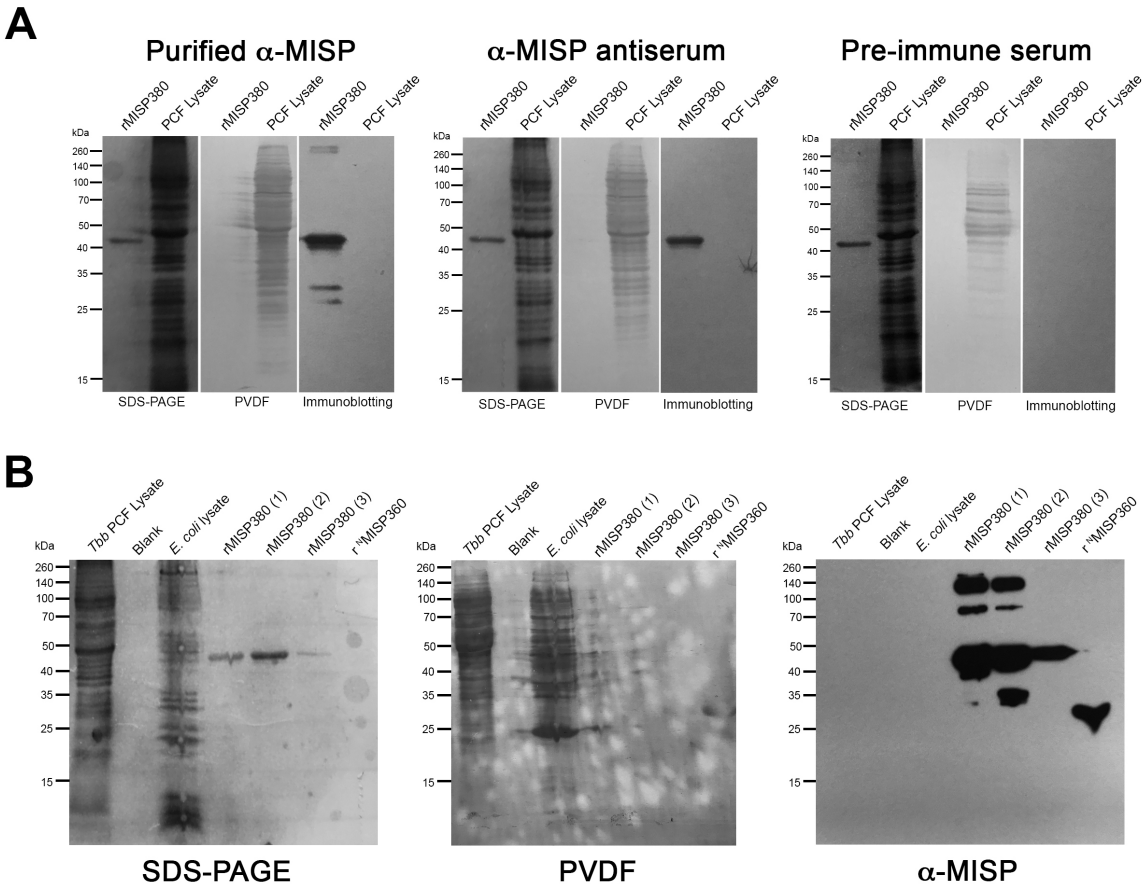


Figure 3.17. anti-MISP polyclonal reactivity tests. (A) Tests using purified rMISP380 and AnTat 1.1. 90:13 wildtype PCF lysate, probed with purified anti-MISP, anti-MISP anti-serum and pre-immune rabbit serum. For each blotting, PVDF

membranes stained with nigrosin and SDS-PAGE stained with InstaBlue are shown as loading control. **(B)** Immunoblotting probing with the purified anti-MISP on PCF lysate, *E. coli* lysate, 3 independently-produced rMISP380 and MISP360 N-terminus recombinant protein.

3.2.5 Gene and protein expression analyses on MISP and BARP

3.2.5.1 Developmental expression of MISP and BARP

The nLC-MS/MS analysis identified MISP peptides in tsetse infected saliva, which originate from parasite stages dwelling in infected SG. To better understand the expression of *misp* during parasite development in the fly, we used semi-quantitative and real time RT-PCR, alongside *barp* detection as a marker for SG EMFs (Urwyler et al., 2007). Total RNA samples were isolated from all *T. b. brucei* life stages obtained from both *in vitro* cultures (cultured bloodstream forms (cBSF) and procyclic cultured forms (PCF)) and fly-derived parasites at 30 d.p.i: i.e. midgut procyclics, proventricular parasites (95% mesocyclics, 5% epimastigotes), whole infected salivary glands (SG) and isolated metacyclic forms (MCF). Samples from naïve flies (30 days old) were used as negative control. Total RNA was reverse-transcribed into cDNA using oligo dT_{20} primers for a faithful representation of mature mRNA. Due to the high sequence identity among the five *misp* homolog genes, assessing the expression of individual isoforms by real-time RT-PCR was not possible. I therefore approached this issue by using the more permissive semi-quantitative RT-PCR method and exploited the differential lengths of the C-termini repeats region, which are unique for 4 of the 5 homologs, and show differences of 78 bp (**Figure 3.18**).

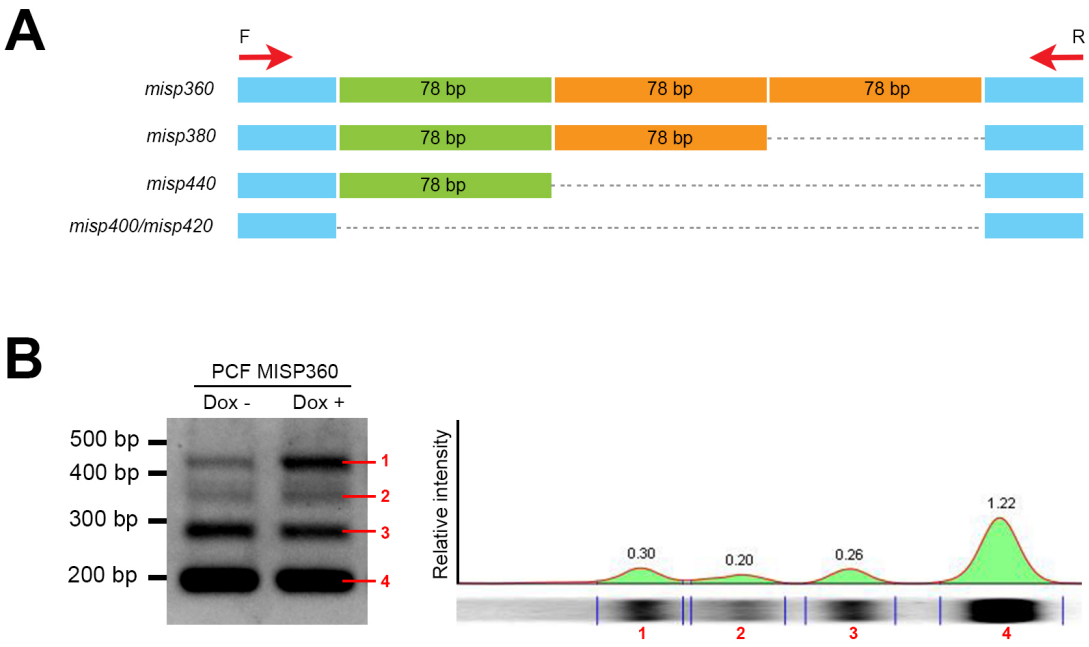


Figure 3.18. Exploiting the *misp* C-termini repeats to determine homolog expression. (A) Schematic of the C-terminus coding region of the 5 *T. brucei* homologs, depicting the 78 bp repeats and the annealing sites of the forward and reverse primers (F, R, red arrows). **(B)** Representative 2% agarose gel displaying the 4 expected bands from RT-PCR reaction products from PCF overexpressing (Dox +) or not (Dox -) *misp360* (left); representative graph for band relative intensity quantification (arbitrary units).

While control samples from naïve flies are all negative for *misp* expression, there is a great up-regulation of *misp* genes in SG stages compared to MG (11.8-fold) or cBSF (25.1-fold), with p-values <0.001 (**Figure 3.19**). Furthermore, although isolated MCF also presented similar expression levels compared to SG (1.5-fold less, p=0.003), we could not rule out that ~5% 'contaminating' EMFs in saliva also contributed to these high expression levels. PV trypanosomes, despite having a 2.1-fold (p<0.001) lower expression compared to SG stages, showed significantly higher expression than MG procyclics, cBSF, and PCF (>5.5-fold, p=0.002) all showing negligible levels of *misp*. Interestingly, although there seems to be an overall preferential expression for *misp* homologs with short C-termini (*misp400/420*), the longer homologs *misp380* and *misp360* are only expressed in SG and MCF. In MCF, the expression of *misp360* represents ca. 25% of the total expression of *misp* genes,

increase that appears to be compensated by the down-regulation of the shortest *misp-400/420.* To further validate the specificity and sensitivity of the detection method, two control cDNA samples were added from a mutant PCF cell line overexpressing an ectopic copy of *misp360* under a tetracycline-inducible system (**Figure 3.20**). The uninduced (Dox-) control cells showed low *misp* expression levels similar to wildtype PCF, while the induced (Dox+) cells showed a specific 2.8-fold increase in *misp*360 expression only, thus proving a good sensitivity and specificity of our detection method (**Figure 3.20**). The expression levels of *barp* (all 14 homologs which share >60% identity at protein level) were also determined using the same samples for comparison (**Figure 3.19**). *barp* transcripts were found to be highly upregulated in infected PV compared to MG forms (10.6-fold), PCF (24-fold) or cBSF (26.6-fold), with p-values <0.01. This high expression levels of *barp* transcripts (-1.2-fold; with p-values of <0.01) is maintained in total SGs (-1.6-fold) and MCF. cBSF, PCF and MG procyclics showed no significant *barp* expression levels. Uninfected tsetse controls were all negative (not shown). To validate these results, a similar expression analysis was performed using real time RT-PCR targeting universal sequences in *misp* and *barp* (**Figure 3.20**). Results showed the same trend in high expression of both *misp* (40-fold) and *barp* (95.6-fold) in SG compared to MG parasites, although with a greater fold change.

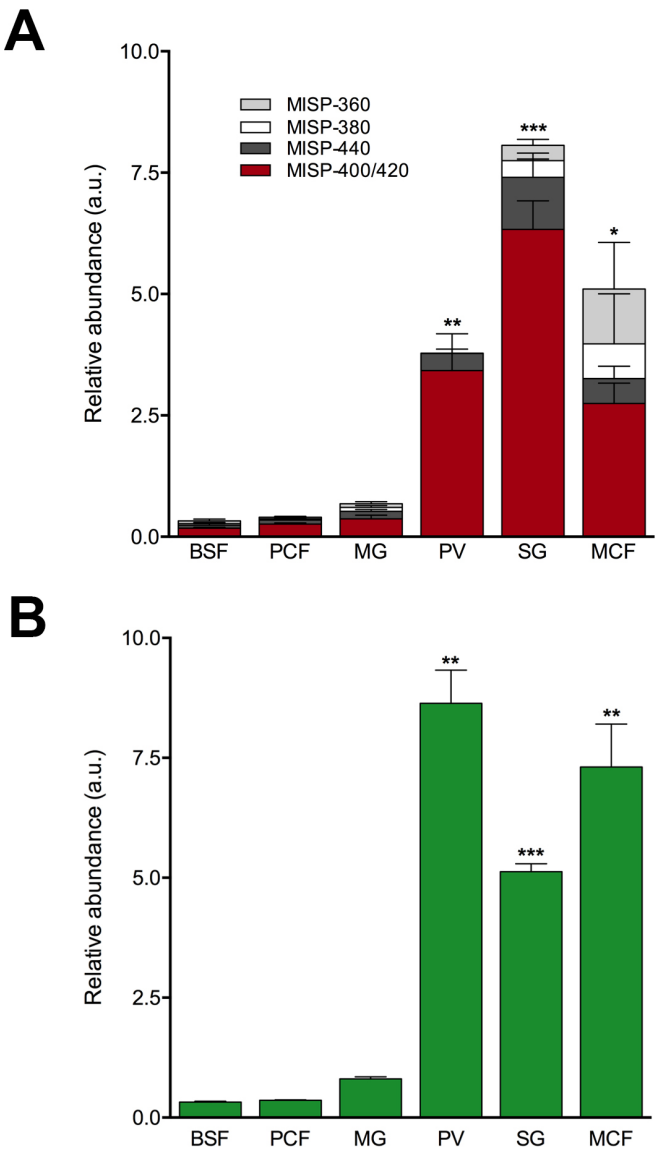


Figure 3.19. Expression levels of *misp* and *barp* throughout the *T. b. brucei* life cycle. (A) Relative mRNA expression levels (arbitrary units) of the *T. b. brucei misp* paralogs, detected by RT-PCR targeting the C-termini motifs in *T. b. brucei* AnTat 1.1 90:13 cultured bloodstream forms (BSF), procyclic cultured forms (PCF), midgut procyclics, proventricular forms, salivary gland forms (SG) and isolated metacyclic forms (MCF). **(B)** Relative mRNA expression levels of *T. b. brucei barp*s, targeting a universal region in all homologs. Error bars represent standard deviation, asterisks show statistical significance (* for p-value <0.05, ** <0.01, *** < 0.001).

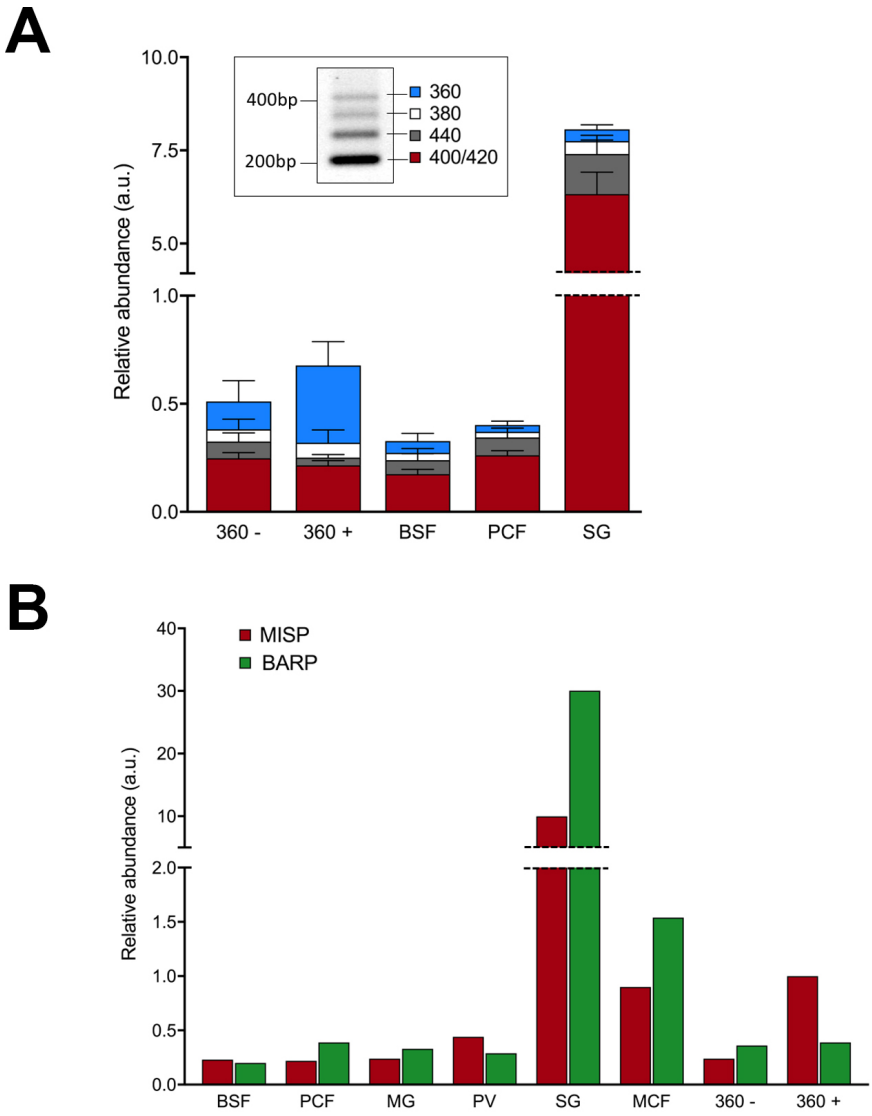


Figure 3.20. Controls of detection method. **(A)** Controls for the semi-quantitative RT-PCR method used to determine the relative RNA expression levels of MISP. BSF, PCF and SG samples shown as in Figure 3.19. The introduced control is the MISP-360 PCF mutant line either uninduced (360 -) or induced (360 +). The inset is a representative image of a DNA agarose gel displaying 4 bands corresponding to the amplification of the different *misp* homologs. **(B)** *misp* and *barp* relative RNA expression levels using quantitative RT-PCR on the samples in Figure 2. Control, same as in panel A.

3.2.5.2 Detecting MISP and BARP proteins in SG parasite stages

Having identified trypanosome SG stages as the major contributors of *misp* and *barp* gene expression, I then determined their expression by immunofluorescence using polyclonal antibodies against MISP and BARP. Neither cBSF or PCF, nor parasites derived from MG or PV were recognised by either antibody (not shown). Recognition only occurred in infected SGs (**Figure 3.21**). Both anti-MISP and anti-BARP antibodies recognised all parasite stages in infected glands, staining clustered cells (EMF and P-MCF) that appeared attached to the epithelium, but also free-swimming MCF. Thus, both MISP and BARP proteins appear specifically expressed in trypanosome forms infecting SGs, and are absent in all other developmental stages, including those in the PV despite the high transcript levels (**Figure 3.19**).

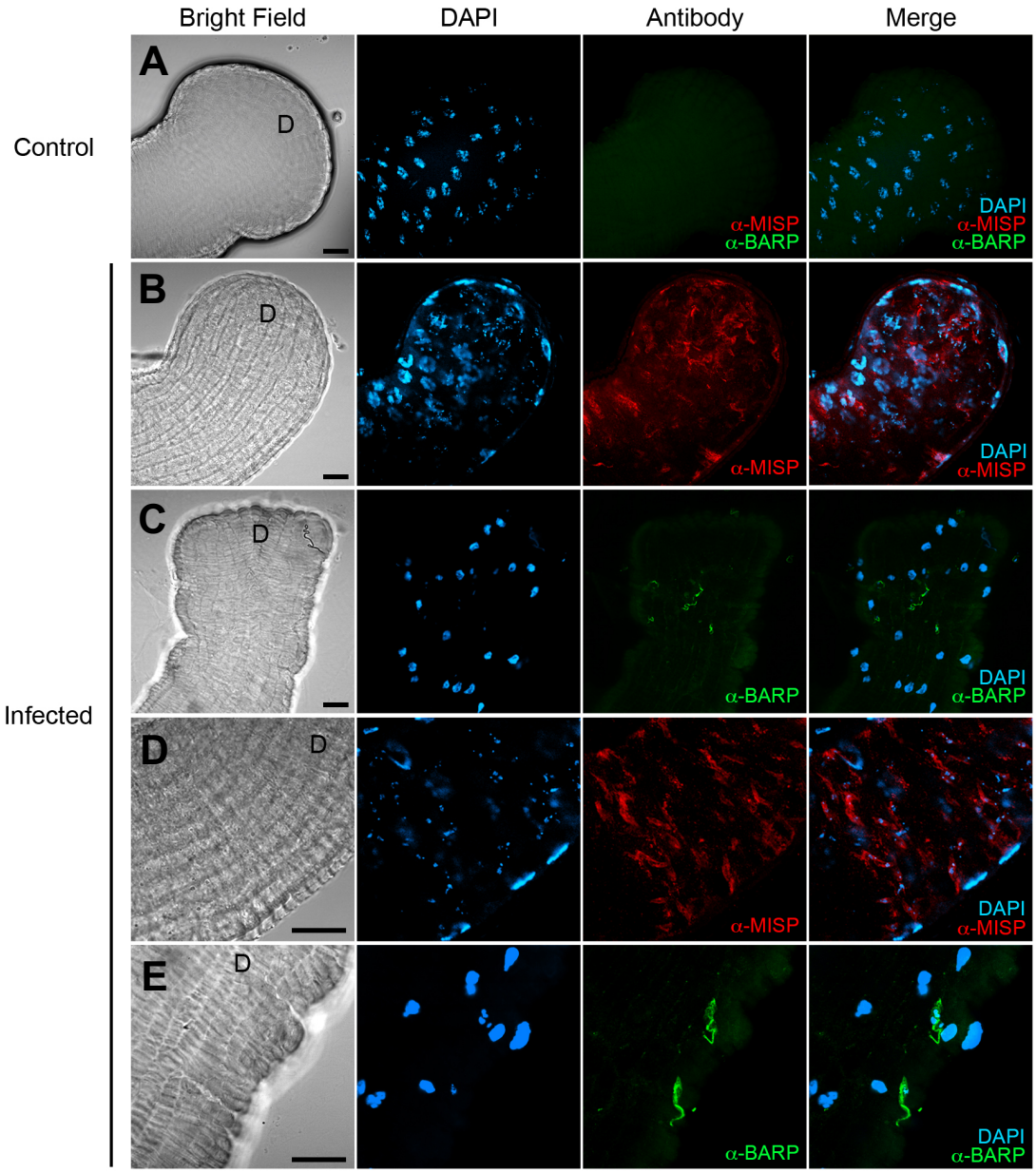


Figure 3.21. *T. b. brucei* salivary gland life stages express MISP and BARP proteins during an infection in the tsetse salivary glands. Immunostaining of MISP and BARP proteins in trypanosomes infecting a tsetse salivary gland ('d' points towards the distal end of the glands). Representative images of uninfected (A) and *T. b. brucei*-infected salivary glands (B-E). Bright field (BF), DAPI (blue), α-MISP (red) and α-BARP (green). Scale bars = 20 μm.

3.2.5.3. Cellular localisation of MISP and BARP

In order to determine the precise cellular localisation of these proteins, I then probed individual PFA-fixed cells extracted from infected SG as well as cBSF, PCF, MG and PV forms (as negative controls). Only EMFs, P-MCFs and MCFs (**Figure 3.22-A**) extracted from SG reacted with anti-MISP antibodies. Relative mean fluorescence intensities showed evidence of a higher recognition in MCFs compared to P-MCF (1.5-fold) and EMFs (2.3-fold) (**Figure 3.22-B**) (p-values 0.01 and <0.001, respectively). This suggests that MISP protein expression begins in EMFs and progresses along with parasite's development within the SG, achieving its maximum level of expression in mature MCFs. MISP appeared to localise evenly throughout the cell surface in non-permeabilised cells (**Figure 3.22-A**). To confirm surface exposure of MISP epitopes (as predicted by molecular modelling analysis, see section 3.2.6.3), immunostaining with anti-MISP was performed on live (instead of PFA-fixed) MCF cells (**Figure 3.23**). In this case, all cells also stained positive and uniformly when incubated with anti-MISP antibody thus suggesting that at least part of MISP N-terminal epitopes are accessible to antibodies. Remarkably, live MCF incubated with anti-MISP at RT showed reduced motility and expansion of the flagellar pocket compared with control cells, although cells recovered when incubated at 37°C (not shown), suggesting they may undergo antibody-mVSG endocytosis like BSFs (Engstler et al., 2007).

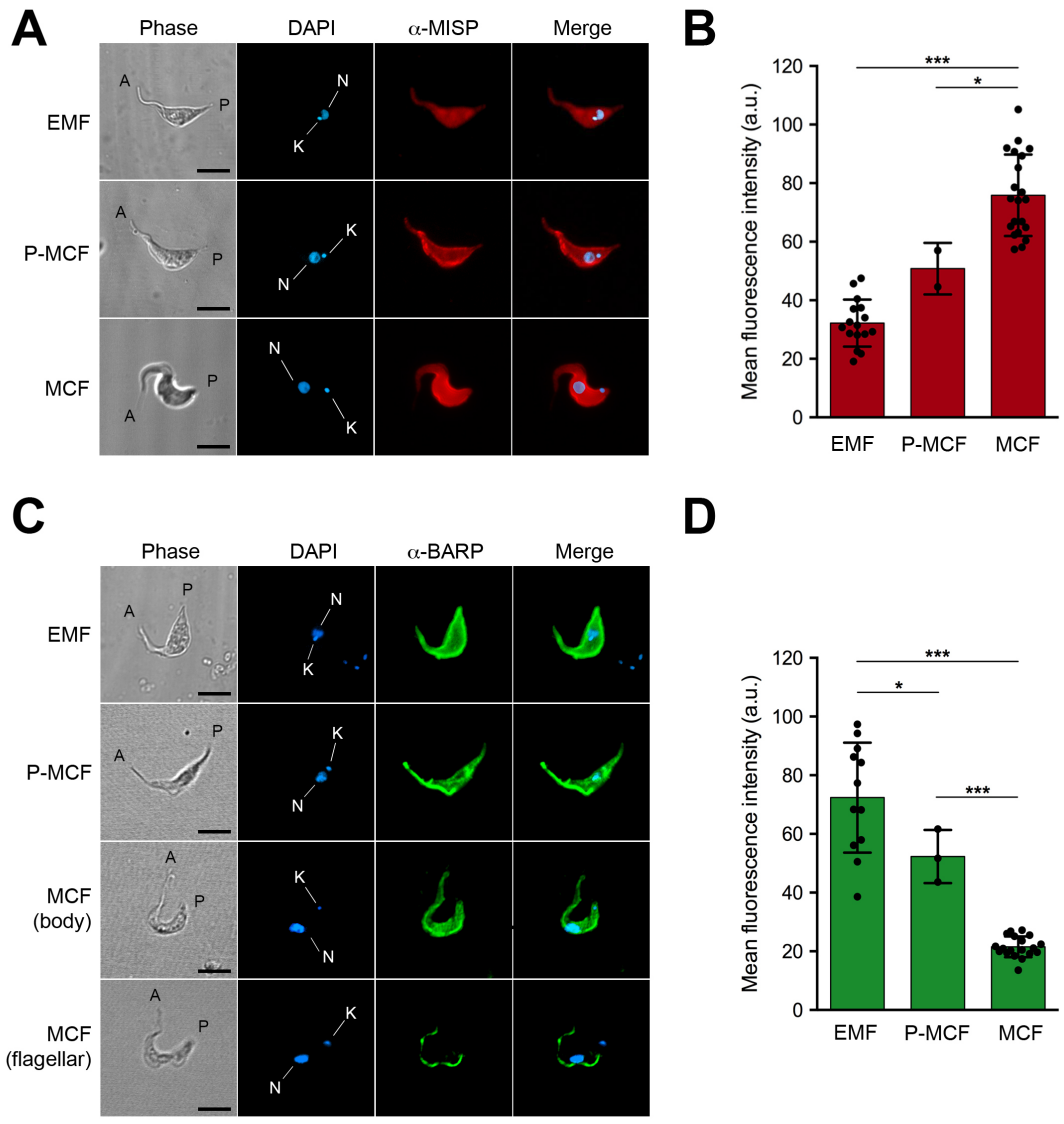


Figure 3.22. Expression and localisation of MISP and BARP proteins in *T. b. brucei* salivary gland stages. (A) Localisation of MISP. Immunostaining of non-permeabilised parasites extracted from infected salivary glands: epimastigote (EMF), pre-metacyclic form (P-MCF) and metacyclic form (MCF). Bright field (BF), antibody α-MISP (red), and DAPI (blue). Kinetoplasts (K) and nuclei (N) are highlighted in the DAPI channel. **(B)** Mean fluorescence intensity (arbitrary units) of anti-MISP stained cells. **(C)** Localisation of BARP. Immunostaining of non-permeabilised salivary gland parasites. Bright field (BF), antibody α-BARP (green) and DAPI. Representative MCFs for BARP flagellar (1) and surface (2) localization. **(D)** Mean fluorescence intensity (arbitrary units) of anti-BARP. Scale bars = 5 µm. Error bars show standard deviation, asterisks represent significance (* for p-value < 0.05; *** for p-value < 0.001).

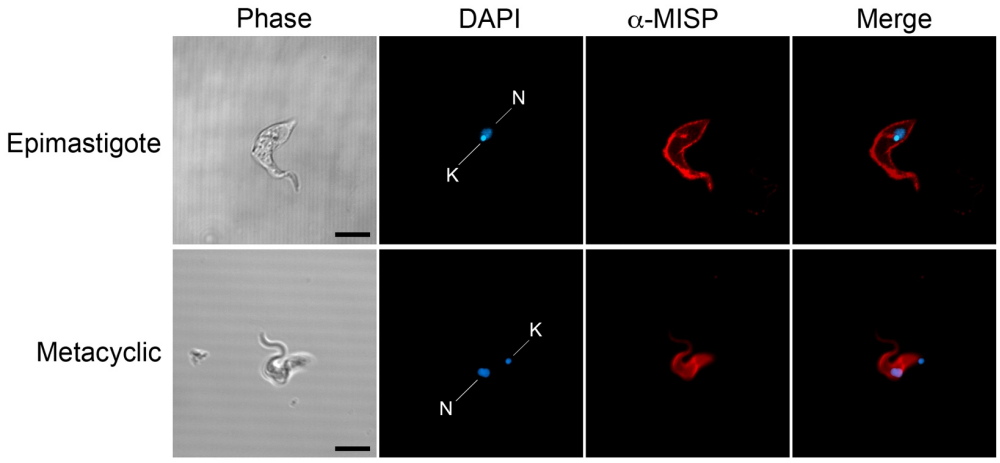


Figure 3.23. Live staining of MCF with anti-MISP. Representative images of immunostaining on live SG parasites using anti-MISP. Epimastigote and metacyclic form stained with anti-MISP (red) and DAPI (blue). Nuclei (N) and kinetoplasts (K) noted in the blue channel. Scale bars = 5 µm.

Interestingly, when an eGFP and HA-epitope tagged MISP360 was ectopically expressed in an induced (Dox+) PCF cell line (**Figure 3.24**), the protein was exclusively seen in the flagellum as suggested by its co-localisation with the flagellar marker (paraflagellar rod, PFR) (**Figure 3.24-C**). However, when the same MISP360 overexpressor cell line progressed through the fly and infected the SGs, the ectopic protein re-localised evenly on the MCF surface (**Figure 3.24-C**), resembling that of wildtype MCF. All non-induced (Dox-) PCF were negative for both anti-HA and anti-MISP antibodies (not shown). I then compared MISP protein expression and localisation in relation to BARP.

Using the same method, BARP was detected on the surface of non-permeabilised EMF, P-MCF and MCF cells (**Figure 3.22-C**). Relative mean fluorescence intensities in EMF were significantly higher compared to P-MCF (1.3-fold) and MCF (3.3-fold) (**Figure 3.22-D**), with p-values of 0.04 and <0.001, respectively. BARP localises evenly throughout the cell surface in EMF and P-MCF cells. However, in MCFs BARP only conserved this localisation in 10.4% of cells (n=481) while the 89.6% of cells showed exclusive flagellar localisation (**Figure 3.22-C**). The expression of BARP in MCF was corroborated using a second anti-BARP polyclonal anti-serum (not shown).

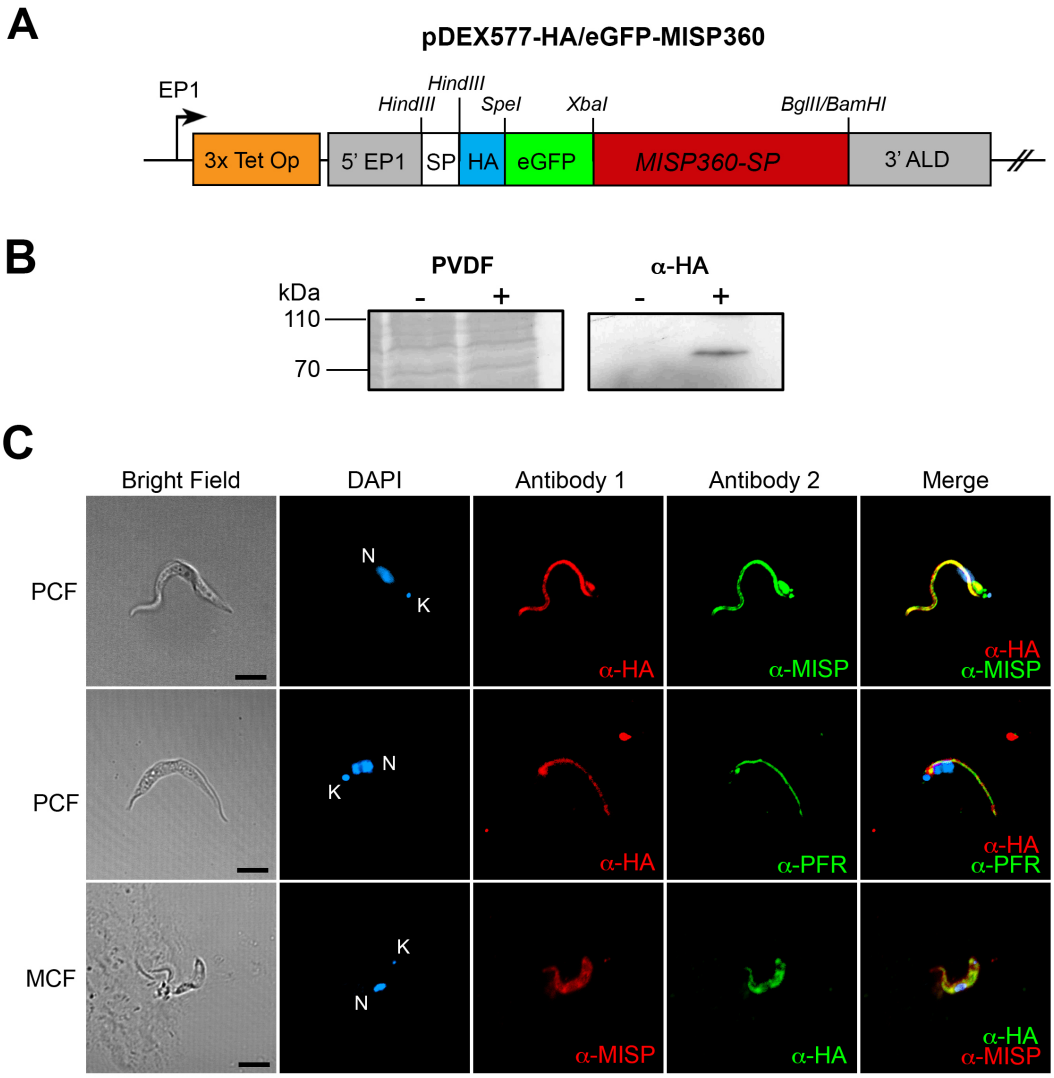


Figure 3.24. Ectopic localisation of tagged MISP360. (A) DNA construct encoding for the ectopic tagged MISP360. **(B)** Immunoblotting to detect ectopic (HA-tagged) MISP360 expressed by the mutant PCF cell line, either with the transgene uninduced (-) or induced (+), probed with anti-HA (top). PVDF membrane stained with nigrosin after film exposure for sample loading control. **(C)** Cellular localisation of the HA-tagged ectopic MISP-360 in AnTat 1.1 90:13 procyclic (PCF) and metacyclic form (MCF), detected using immunostaining on fixed non-permeabilised PCF cells with either anti-HA plus anti-MISP (co-localisation of ectopic tag with MISP), or with anti-HA plus anti-PFR (flagellar localisation) in methanol-fixed cells. Scale bars = 5 μm.

3.2.5.4 Indirect detection of mVSGs in SG parasites

To confirm identity of MCFs from infected SGs, cells were probed with anti-CRD antibodies (Zamze et al., 1988), which recognise the cross-reacting determinant epitope. In the absence of specific anti-VSG antibodies expressed by MCFs, anti-CRD antibodies have been successfully used by others to identify this parasite stage (Rotureau et al., 2012). The CRD epitope is formed when GPI-anchored surface proteins, lacking an acyl fatty acid chain on the inositol ring, are exposed to GPI-specific PLC (GPI-PLC), which cleaves the phosphodiester bond between the inositol ring and the lipid moiety, forming a 1,2 cyclic phosphate ring on the inositol residue (i.e. CRD epitope) (Zamze et al., 1988). This enzyme cleaves the GPI detaching the protein from the membrane. For example, the di-myristoyl-PI anchor of the *T. brucei* "membrane-bound" mVSG is completely susceptible to the parasite GPI-PLC forming "soluble" sVSGs (Cardoso de Almeida and Turner, 1983; Zamze et al., 1988). Monomeric proteins with a single GPI anchor become soluble after cleavage, while dimeric proteins, such as VSG, may be only cleaved at one lipid anchor (thus exposing the CRD epitope in only one GPI glycan) and remain attached to the membrane by the second one. Because of the presence of VSGs on the surface but also due to the developmental expression of the GPI-PLC (Christiano et al., 2017), the BSF and MCF stages are the only forms of *T. brucei* reactive to anti-CRD. As shown in **Figure 3.25**, neither attached EMF nor P-MCF react to the antibody, and only MCF display positive signal on the surface as expected. This confirms the SG stages identified for the expression and localisation of MISP and BARP based on morphology, size and relative position of the kDNA. Technical reasons prevented unfortunately the co-localisation of CRD epitopes and MISP or BARP because all three antibodies were made in the same animal host.

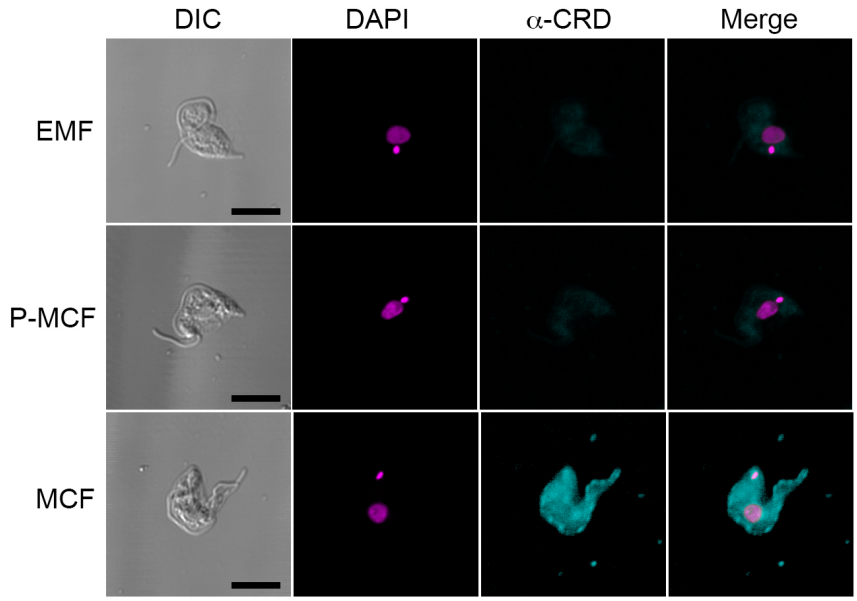


Figure 3.25. Indirect detection of mVSG in *T. brucei* SG stages. Immunostaining of EMF (AE), pre-metacyclic (P-MCF) and metacyclic (MCF) stages from tsetse infected SG with anti-CRD antibodies. Differential interference contrast (DIC), DAPI (magenta) and anti-CRD (blue). Scale bars = 5 μm.

3.2.6 Structural studies on MISP and BARP

3.2.6.1 The crystal structure of MISP N-terminus

The conserved N-terminal domain of MISP360 (Tb427.07.360, Gly^{24}–Ala^{234}) was recombinantly produced in *E. coli* (rN-MISP360) and purified to homogeneity using nickel affinity and size exclusion chromatography (SEC) (**Figure 3.26-A**). Comparison of the SEC elution profile against a series of globular protein standards showed that rN-MISP360 eluted as a monomer of approximately 22 kDa (**Figure 3.26-B**). Crystals of purified rN-MISP360 were obtained using the hanging drop method and the structure of rN-MISP360 was determined by molecular replacement using a truncated form from *T. congolense* GARP (PDB: 2y44) as the search model. The identification of *Tc*GARP as a suitable model was based on secondary structure predictions as the aminoacidic sequence identity of the N-terminal ectodomain is only 15% (see section 3.2.2.4). The overall structure of rN-MISP360 was refined to a resolution of 1.82 Å (structure obtained by Prof. Martin Boulanger, University of Victoria). The core of the N-terminal ectodomain adopts an elongated structure

measuring approximately 83 Å in height and spanning approximately 25 Å in width (**Figure 3.26-C**). The ectodomain is well ordered with low *B*-factors throughout the protein (**Figure 3.26-D**). Similar to *Tc*GARP, rN-MISP-360 adopts an overall triple helical bundle structure composed of a core of extended twisted helices capped by a smaller helical bundle at the N-terminal.

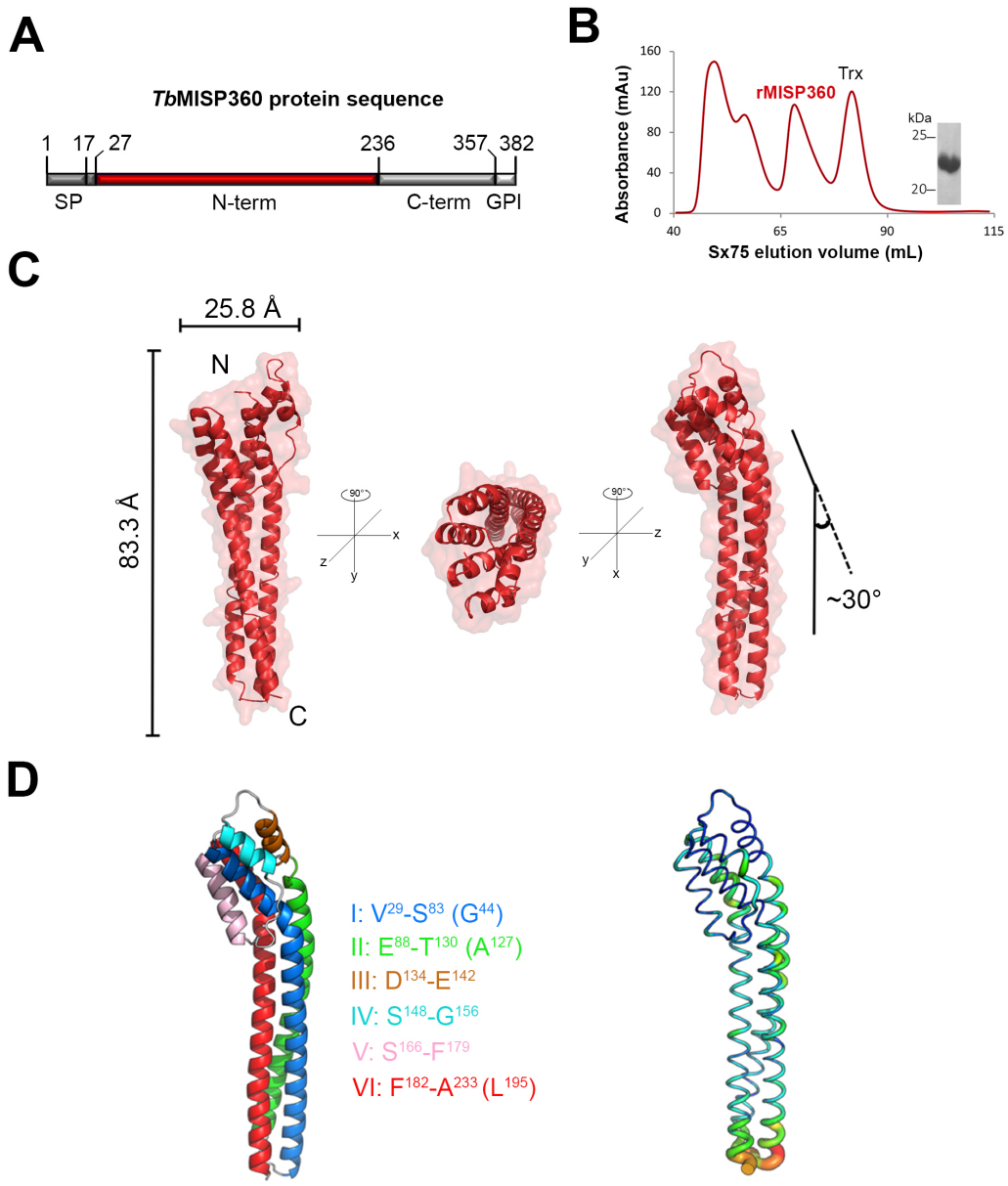


Figure 3.26. The crystal structure of the MISP360 N-terminus. (A) Construct encoding the MISP360 N-terminus that was recombinantly expressed in *E. coli* and crystallized (N24-E235). **(B)** Superdex 75 column size exclusion chromatogram of recombinant MISP360 and SDS-PAGE analysis of the column fraction (inset) with the

protein migrating at ~22 kDa. **(C)** MISP360 N-terminus crystal structure with surface representation. The structure was found to be 83.3 Å tall, 25.5 Å in width with a 30° helix bend at the top. **(D)** Secondary structure depiction highlighting the organization of the helices. The bend-forming residues are in parentheses: helix I (G44), helix II (A127) and helix VI (L194). Coiled structures are represented in *white* (left); *B*-factor putty model with ordered regions displayed in blue and thin tubes and flexible regions in red and thicker tubes (right).

The helical bundle that dominates the structure consists principally of three helices (**Figure 3.26-D**), which adopt a bend of approximately 30° and, collectively, give rise to the helical bundle cap. In addition to the ends of the 3 major helices, this bundle includes 3 shorter helices. Moreover, the head structure of rN-MISP-360 is anchored by two disulphide bonds: one between Cys^{36} and Cys^{157}, and the other between Cys^{177} and Cys^{185}.

3.2.6.2 BARP adopts a triple helical bundle structure

Attempts were also made to determine the crystal structure of BARP. A recombinant protein of BARP (Tb927.9.15630) was produced and crystallized, but the resultant structure had too low resolution (3.9 Å). However, using this structure and the crystal structures of MISP360 and *Tc*GARP (PDB: 2Y44), a high confidence model of BARP was built (**Figure 3.27**) using Modeller 9v18 (Webb, 2016) (Prof. Martin Boulanger, University of Victoria). BARP adopts an almost identical triple helical bundle structure that of MISP. Unlike MISP, BARP does not have a C-terminal domain to include in the model (**Figure 3.28-B**).

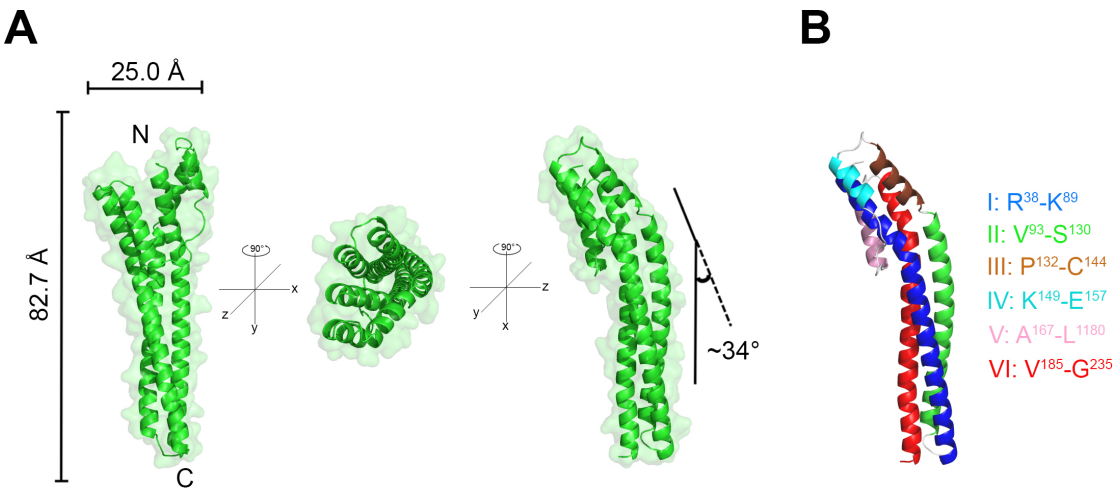


Figure 3.27. High confidence model of *Tb*BARP (Tb927.9.15630). (A) Cartoon representation of the model, in three different angle views. The structure is overall 82.7 Å tall, 25.0 in width and a 34° helix bend. **(B)** Cartoon representation depicting the helices organisation.

3.2.6.3 Modelling the MISP C-termini

The C-terminus domain of MISP360 (Lys^{238}–Ser^{357}) was not included in the crystallised rN-MISP360 as it is less conserved across isoforms and it is predicted to lack defined secondary structures and to adopt random coiled dispositions (see section 3.2.2.4). As shown in **Figure 3.28-A**, the N-terminal domains of MISP, BARP and VSG (PDB: 1VSG, (Freymann et al., 1990) are very similar in height (MISP 6.9 Å and BARP 10.7 Å shorter than VSG). However, their respective C-terminal domains are variable and could result in very different spatial arrangements on the MCF surface. To understand how these proteins are displayed and exposed on the MCF surface, a sequence length comparison was done between the *T. brucei* MISP C-termini, two VSG C-termini whose structure has been solved and 2 BARPs (**Figure 3.28**). All BARP isoforms have no C-terminal domain except for Tb927.9.15590 that has a C- domain slightly shorter than VSGs. Both VSGs (VSG MiTat 1.1 and MiTat 1.2) have equivalent C-terminal lengths, comparable to that of MISP440. While MISP400 and MISP420 have shorter domains compared to VSGs, MISP360 and MISP380 have a ~71% and ~35% longer C-termini than VSGs respectively. Using the I-TASSER and IntFOLD severs, the C-termini of all *T. brucei* MISP isoforms were modelled and attached to their corresponding N-termini domains (**Figure 3.29**). These

were then modelled in the MCF surface context attached to the plasma membrane via GPI-anchor and surrounded by mVSGs. As predicted from the sequence length comparison, the MISP isoforms seem to adopt variable levels of exposure depending on the C-terminal domain length, which is a result of the number of repeats (see **Figure 3.8-B**). Thus, MISP360 is potentially the most exposed isoform, followed by MISP380, MISP440 and MISP400/420 (in that order). However, these C-termini structural models are of low confidence and it is unlikely they adopt these dispositions rather than folding and contracting in unpredictable ways. Thus, based on length, MISP400/420 are likely to be masked by the mVSGs, whereas MISP440 would be partly protected. In contrast, both MISP380 and MISP360 are likely to be exposed over the VSG coat. As shown in **Figure 3.23**, at least one MISP isoform is indeed exposed and available for antibody binding, supporting this model. However, it may be possible that when co-expressed with BARP in EMFs, MISP are exposed to the exterior regardless of the isoform expressed.

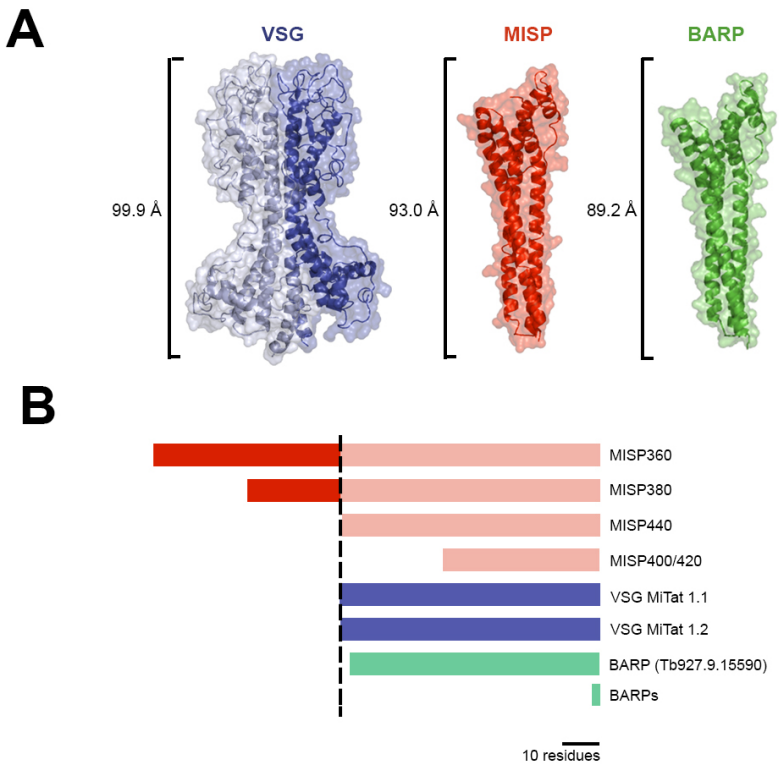


Figure 3.28. Comparison between MISP and VSG C-termini. (A) Structural comparison of MISP and VSG N-termini with their respective height. **(B)** Schematic length comparison between the five *T. brucei* MISP isoforms and two VSGs structurally characterised.

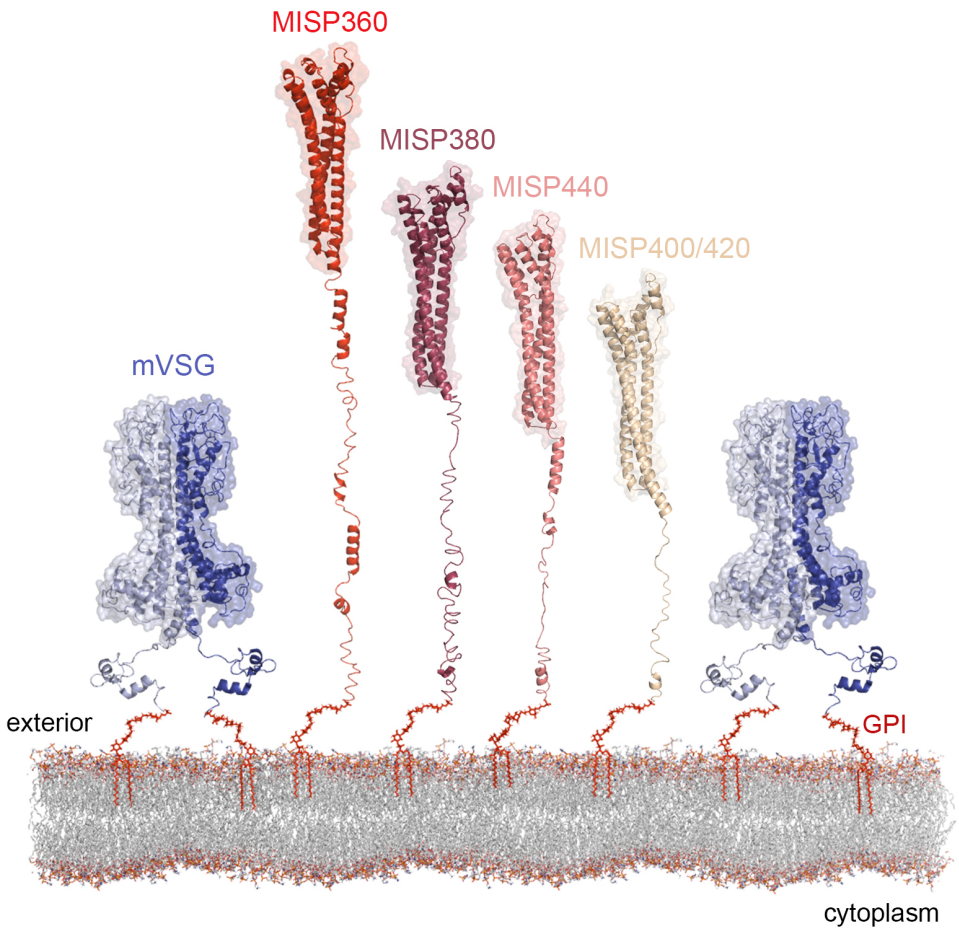


Figure 3.29. Proposed surface display of all MISP isoforms when expressed by *T. brucei* MCFs. The model of this metacyclic plasma membrane contains m-VSGs and displays the MISP isoforms modelled from the MISP360 N-terminus crystal structure using I-TASSER.

3.2.6.4 Structural homologs of MISP and BARP

The lack of significant sequence identity between MISP-360 and any protein with known function led us to perform a DALI (Holm and Rosenstrom, 2010) search to identify structural homologs. *Tc*GARP was identified as the top hit with a Z-score of 21.4. The DALI also revealed structural homology to the previously characterised haptoglobin-haemoglobin receptor (HpHbR) from *T. brucei* (PDB: 4x0j) (Lane-Serff et al., 2014)and *T. congolense* (PDB: 4e40) (Higgins et al., 2013; Lane-Serff et al., 2016; Stodkilde et al., 2014), and to a *T. brucei* variant surface glycoprotein monomer (VSG) (PDB: 1VSG) (Freymann et al., 1990) with Z-scores of 18.1, 17.7 and 6.5,

respectively. All these proteins exhibit a complementary core of twisted three helical bundles (**Figure 3.30**). Structural comparison, however, indicates a closer architectural similarity with HpHbR compared to VSG. This is due, in part, to the breakdown of third helix into loops and extensions enabling substantial conformational diversity (Freymann et al.). Moreover, rN-MISP-360 is monomeric in contrast to dimeric VSGs (Freymann et al.). Despite the general structural similarity with *Tc*GARP, *Tb/Tc*HpHbR and the VSG monomer, the overall low sequence identity and the lack of key conserved residues across them indicates a likely different biological role for MISP, as happens with GARP (unknown), *Tb/Tc*HpHb (haeme binding) and VSG (antigenic variation).

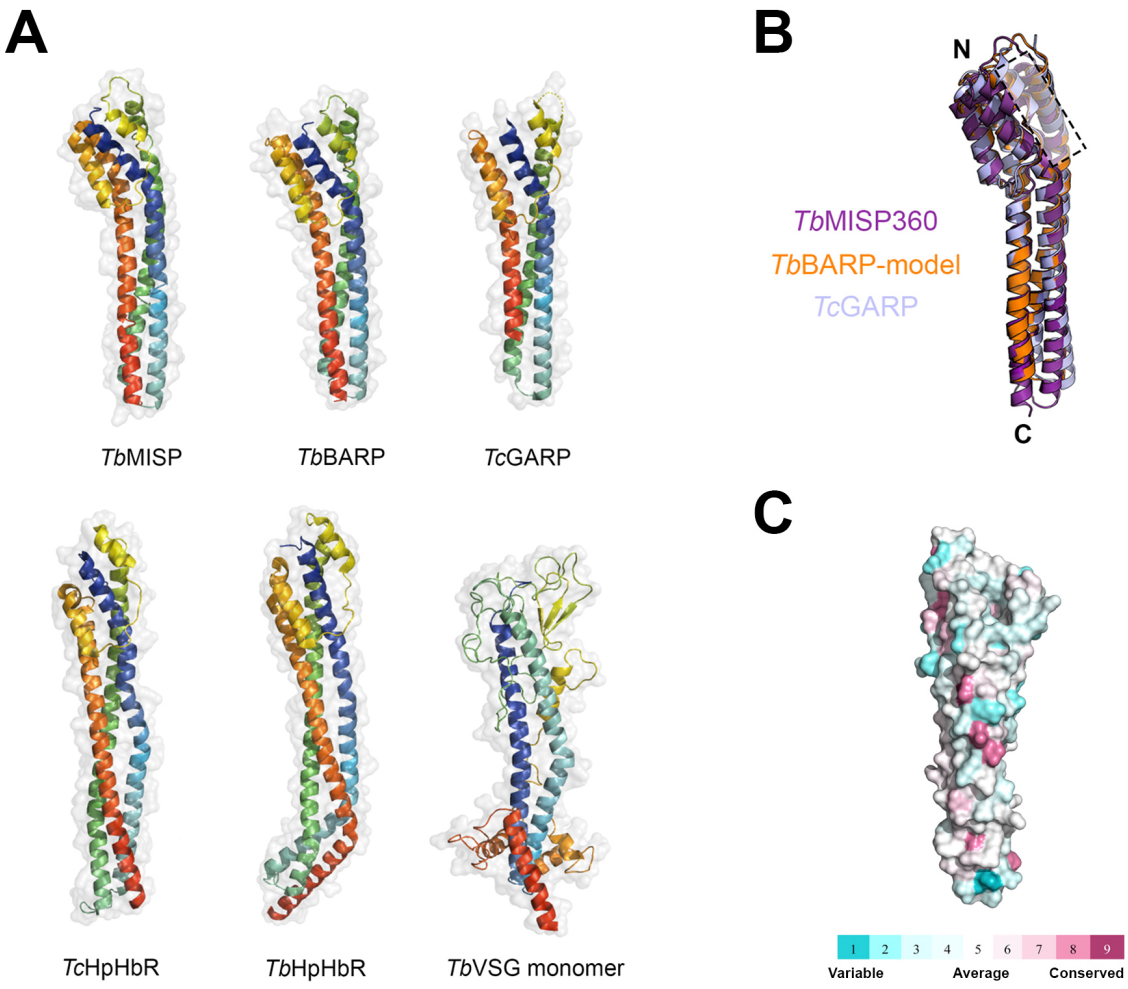


Figure 3.30. MISP structural homologs. (A) Structural comparison of the top-scored structural homologs of MISP in DALI search, coloured from blue (N-termini) to red (C-termini). **(B)** Structural superimposition of MISP, BARP (high confidence model), and

GARP. **(C)** Residue conservation between MISP, BARP and GARP on the crystal structure of MISP. Colour code from blue (most variable) to purple (most conserved).

3.2.6.5 MISP and BARP display a surface molecular pocket

A close analysis to rN-MISP360 and BARP structures revealed a similar distribution of acidic and basic patches along the entire length of the structure and no clear localised charge densities that would indicate a molecular recognition site. However, both structures present a surface pocket at the membrane distal end near the region where the core helices bend in both the proteins (**Figure 3.31**). In rMISP360, the pocket is formed by the N-terminal portion of helix I, the loop connecting helix II to III, helix V and N-terminal region of helix VI. The secondary structures contributing to pocket formation in *Tc*GARP were similar with the N-terminal portion of helix I, C-terminal region of helix II, the loop connecting helix III to IV, and N-terminal region of helix V forming the pocket. Intriguingly, the overall dimensions of the pockets are quite different. For instance, the measure of the rMISP360 pocket is approximately 12.2 Å in depth and 5.9 Å in diameter (**Figure 3.31**), and is lined by ten hydrophobic residues (Ala^{35}, Leu^{37}, Phe^{38}, Leu^{42}, Ile^{133}, Leu^{169} Ala^{172}, Gly^{176}, Ile^{184}, and Leu^{188}). In contrast, the pocket of BARP is significantly smaller measuring approximately 5.7 Å in depth and 6.4 Å in diameter and formed by predominantly polar residues (Thr^{46}, Asp^{49}, Val^{50}, Gln^{53}, Thr^{137}, Glu^{168}, Thr^{170}, Ser^{171}, Ala^{177}, Leu^{181} and Asp^{190}). Analysis of the solvent networks in these pockets revealed that the waters occupied the periphery of the pocket in both *Tc*GARP and rMISP360 with 5 and 3 water molecules, respectively. Importantly, both structures of BARP and rMISP360 lack a short peptide at the N-terminus (QAGDGEDCGGQSVPPK and DSIIEEGS respectively), which are not predicted to fold on the protein but may potentially alter the pocket structures and accessibility to ligands or solvent. Mapping of conserved residues on rMISP360 using ConSurf (**Figure 3.31**) (Ashkenazy et al., 2016; Ashkenazy et al., 2010), revealed that most of the conserved domains are found forming the core helices and the pocket core. However, a striking 51% of the non-conserved residues (which represent a 22.3% of the total sequence) localized right around the pocket, suggesting these residues may mediate affinity or recognition of the pocket's putative ligand.

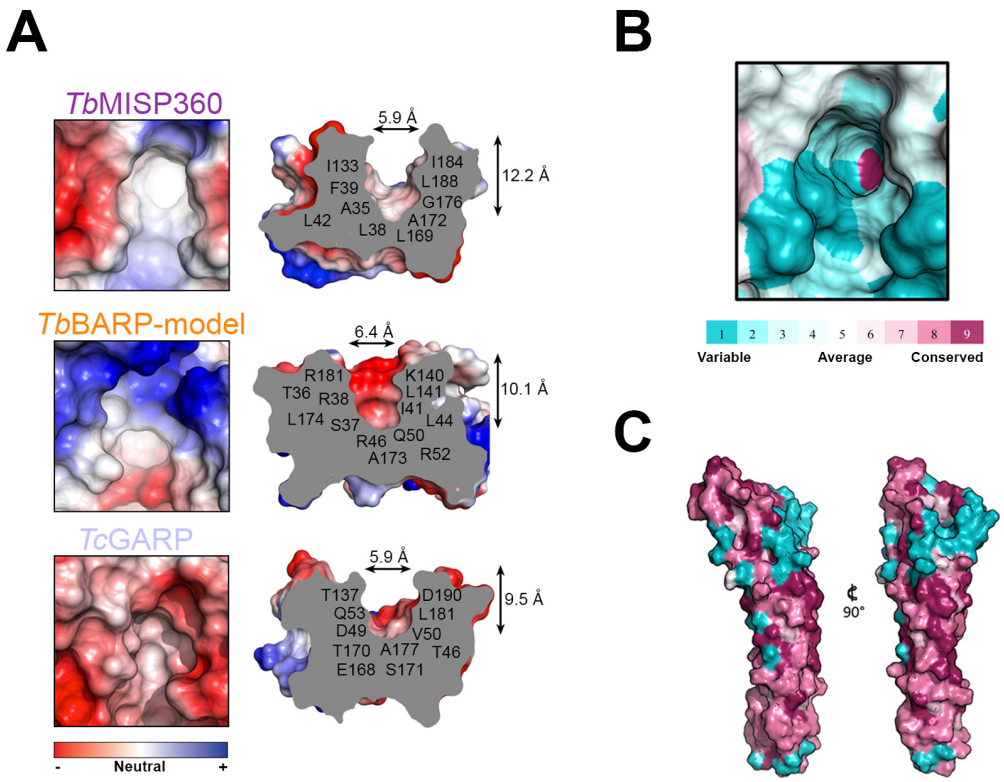


Figure 3.31. MISP surface pocket and residue conservation. (A) Top view (left) and side view (right) of the surface molecular pockets of *Tb*MISP360 (crystal structure), *Tb*BARP (high confidence model) and *Tc*GARP (PDB: 2y44). Residues coloured from red (negatively charged) to white (neutral) and blue (positively charged). Side views show the depth, width and forming residues of the pockets. **(B)** Residue conservation between pocket forming residues of *Tb*MISP, *Tb*BARP and *Tc*GARP on the MISP pocket structure. **(C)** Residue conservation between the five MISP isoforms on the MISP360 crystal structure.

3.2.6.6 Surface exposure of MISP suggests that MCFs express a more permissive VSG coat

From the crystal structure of rMISP360 and the C-termini sequences, structural models for all MISP isoforms were built using I-TASSER and IntFOLD (**Figure 3.32**). Considering MISP as an outer surface glycoprotein attached to the membrane via a GPI anchor, MISP are likely to point their C-termini towards the cell membrane and to project the N-termini towards the exterior. Similar to *Tc*HpHb or a

VSG monomer, it was predicted that the MISP monomeric triple helical bundle is vertically oriented and, within the metacyclic surface context, in close contact with neighbour mVSG homodimers. However, since MISP have longer C-terminal sequences than any known C-terminus VSG type, and despite their lack of secondary structure, we predict that MISP N-termini are likely to reach higher levels of exposure than the mVSG coat (**Figure 3.32**). This is partly supported by the results obtained in MCF immunostaining, where the anti-MISP polyclonal antibody was able to specifically bind to MISP on live metacyclic cells (**Figure 3.25**). Together with the results of immunostaining experiments with anti-BARP, these data suggest that the surface of MCF is not only composed by a dense coat of mVSG homodimers, but instead allows the co-expression of invariant GPI-anchored glycoproteins, like BARP and MISP, the latter more abundant and completely accessible to antibodies (**Figure 3.32**).

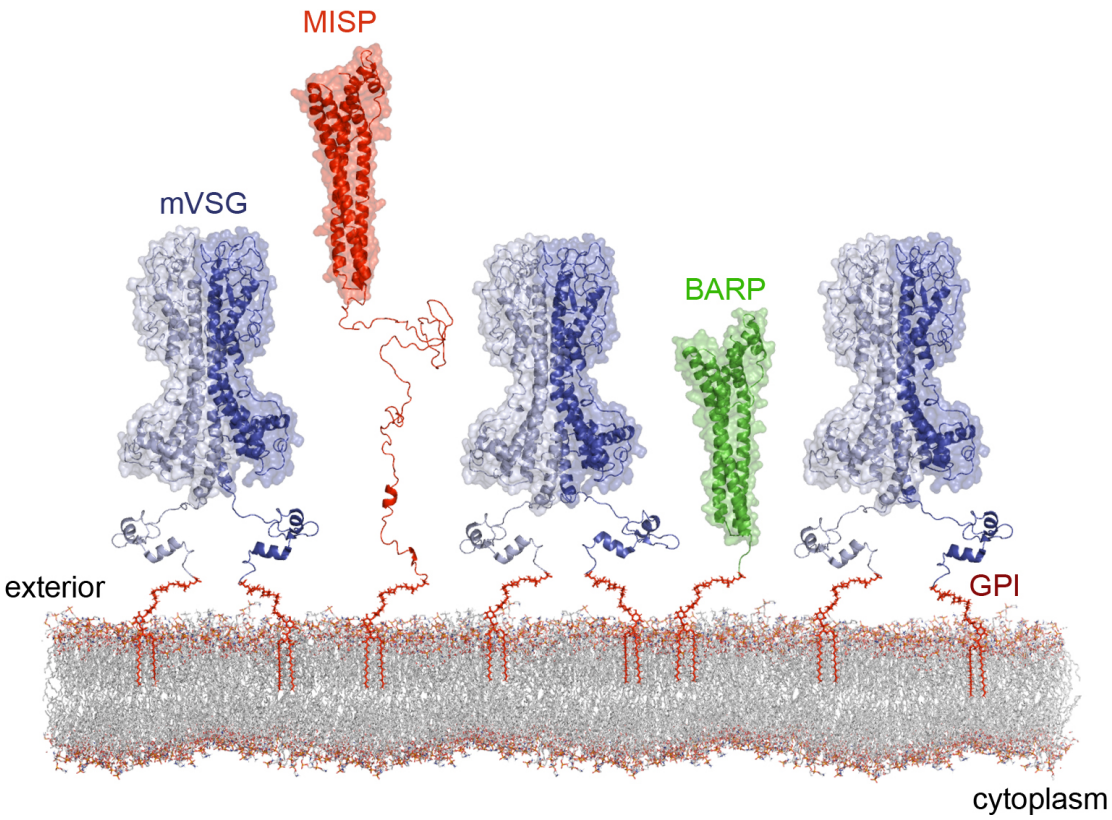


Figure 3.32. Structural model of the metacyclic surface coat. Model of the *T. brucei* metacyclic surface glycocalyx displaying the GPI-anchored VSG homodimers (light and deep blue), MISP (red) and remains of BARP. The mVSG structure is

represented with a low confidence model of mVAT4 based on the crystal structure of VSG 221 N- (PDB: 1VSG) and C-terminus (PDB: 1XU6); MISP is represented with the crystal structure of MISP360 N-terminus (PDB: 5VTL) and a model of its C-terminus; BARP represented with a high confidence model.

3.2.6.7 Re-defining the *T. brucei* life stage-specific glycocalyx

During its life cycle, *T. brucei* constantly keeps differentiating into multiple life stages to overcome external challenges (Matthews, 2005). During many of these changes, the parasite replaces its surface coat primarily composed by GPI-anchored glycoproteins (Ferguson, 1997). The discovery of BARP (previously known as the major surface glycoprotein on SG EMFs, (Urwyler et al., 2007) and MISP on the MCF surface will re-define our understanding of the changes in composition of the *T. brucei* surface glycocalyces during differentiation (Figure 3.32).

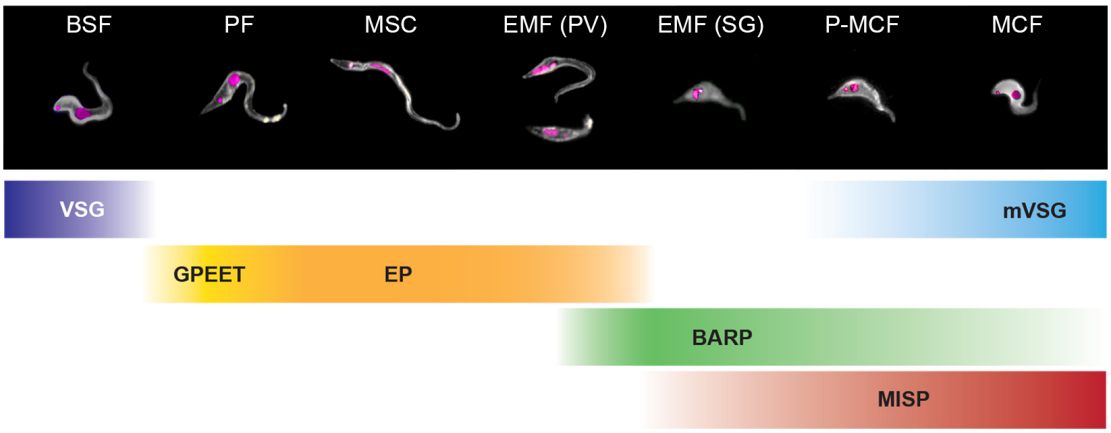


Figure 3.33. The *Trypanosoma brucei* sequential expression of major surface GPI-anchored glycoproteins during its life cycle. Bloodstream forms (BSF), procyclic forms (PF), mesocyclic forms (MSC), short (SE) and long (LE) epimastigotes infecting the proventriculus (EMF PV), attached epimastigotes infecting the salivary glands (EMF SG), pre-metacyclic forms (P-MCF), metacyclic forms (MCF). Representative immunostaining images of the parasite stages highlighting the nuclei and kinetoplasts (magenta), and the cell surface coloured according to the major surface glycoprotein expressed (top). Bottom bars define the duration of the expression of the surface proteins in relation to parasite developmental stage.

3.2.7 Functional studies on MISP

3.2.7.1 Determining the essentiality of MISP during a tsetse infection

To gain more insight into the function of MISP, I created a PCF mutant cell line (*misp*^{RNAi}) to knockdown all *misp* through a tetracycline-inducible RNAi stem loop system (**Figure 3.34-A**) (MacGregor et al., 2013). As expected, based on the low transcript levels, silencing of *misp* in PCF did not show any growth or morphology phenotype *in vitro* (**Figure 3.34-B**). The essentiality of MISP in the tsetse was then analysed with the *misp*^{RNAi} mutant. Trypanosomes were isolated (30 d.p.i) from MG, PV and SG to determine *misp* knockdown levels (**Figure 3.34-C**). In both PCF and MG procyclics, which express negligible levels of *misp*, the transcript levels in induced (Dox+) cells were slightly upregulated (19%; p=0.04) and downregulated (5.9%; p=0.56), respectively, compared to the control. In contrast, when *misp* expression is switched on, we observed a transcript reduction of 58.2% in PV (p-value=0.05) and 31.5% in SG trypanosomes (pooled samples from three replicates). Representative samples of MG, PV and SG were also processed for immunostaining to indirectly quantify the reduction in MISP expression using the polyclonal anti-MISP antibody. As in the wildtype, either induced and non-induced MG and PV forms were negative for staining. However, as shown above, SG EMF and MCF reacted with the antibody (**Figure 3.34-D**) and Dox+ MCFs showed a significant 63.1% reduction in mean fluorescence intensity (p-value <0.001) (**Figure 3.34-E**). No defects in cell shape or motility were observed. Importantly, despite the reduction of *misp* in Dox+ cells, no significant alterations in the infection rates of MG, PV and SGs were observed compared to the non-induced (Dox-) control group (**Figure 3.34-F**).

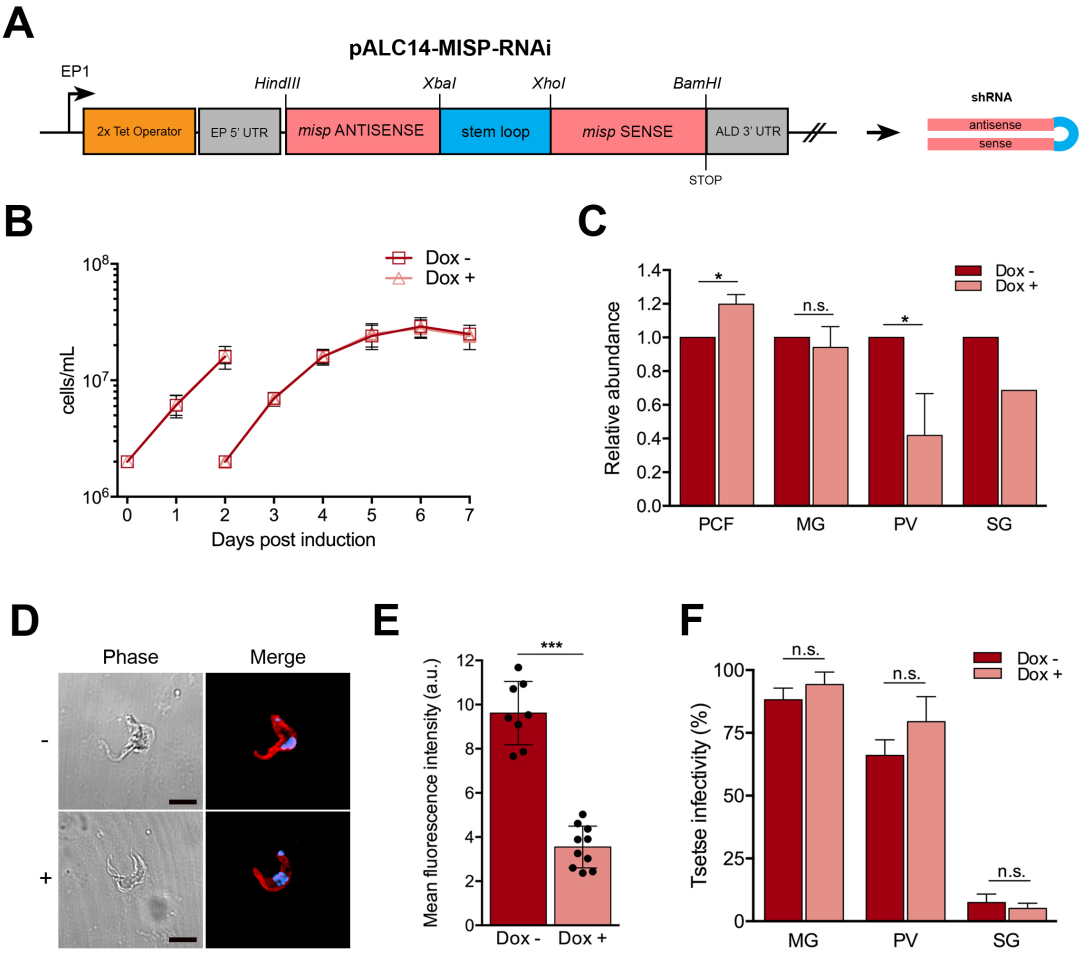


Figure 3.34. ***T. b. brucei misp*** **do not appear to be essential during parasite development in the tsetse. (A)** DNA construct engineered for the tetracycline-inducible expression of a shRNA targeting all *misp* transcripts. **(B)** *In vitro* growth curve of *misp*^{RNAi} PCF cells, either induced (Dox+) or uninduced (Dox-). **(C)** Relative *misp* RNA expression in *misp*^{RNAi} PCF cells, midgut procyclics, proventricular parasites and salivary gland forms (SG). Expression levels of Dox- cells are normalized to 100%. **(D)** Representative immunostaining images of non-permeabilized *misp*^{RNAi} MCF (Dox-/+) with anti-MISP (red), DAPI (blue) and phase. **(E)** MISP mean fluorescence intensities (arbitrary units) on *misp*^{RNAi} MCF (Dox -/+). **(F)** Percentage of flies with *misp*^{RNAi} cells infecting the midgut, proventriculus and salivary glands (SG). Error bars show standard deviation, asterisks represent significance (* for p-value < 0.05; *** for p-value < 0.001). Scale bars = 5 μm.

3.2.7.2 Determining putative binding partners of MISP

Having found that MISP seem not to be essential for *T. brucei* development in the tsetse and considering that MISP function cannot be inferred from homology searches, I next tried to identify potential MISP binding ligands. A recombinant MISP400 was produced in *E. coli* lacking the signal peptide and fused at C-terminus to a biotinylation target sequence. The recombinant biotinylated protein was used in Biacore resonance and pulldown assays by incubating the protein with either mammalian serum or tsetse saliva, the two environments where MCF cells can be in. In both assays, a recombinant *T. brucei* HpHbR, known to have a similar structure to MISP and to bind to haemoglobin (Lane-Serff et al., 2014), was used as negative control. Biacore results showed that MISP400 seems not to bind to any component from neither tsetse saliva (uninfected and infected) or human or bovine serum, while the HpHbR control only showed strong binding to both human and animal sera (not shown). In pulldown experiments, after samples were run in silver-stained SDS-PAGE gels, ~7 unique protein bands were revealed resulting from the interaction of MISP with proteins from both infected and naïve saliva (**Figure 3.35**). Although further identification by mass spectrometry needs to be carried out, these results suggest that MISP400 may interact with tsetse salivary proteins.

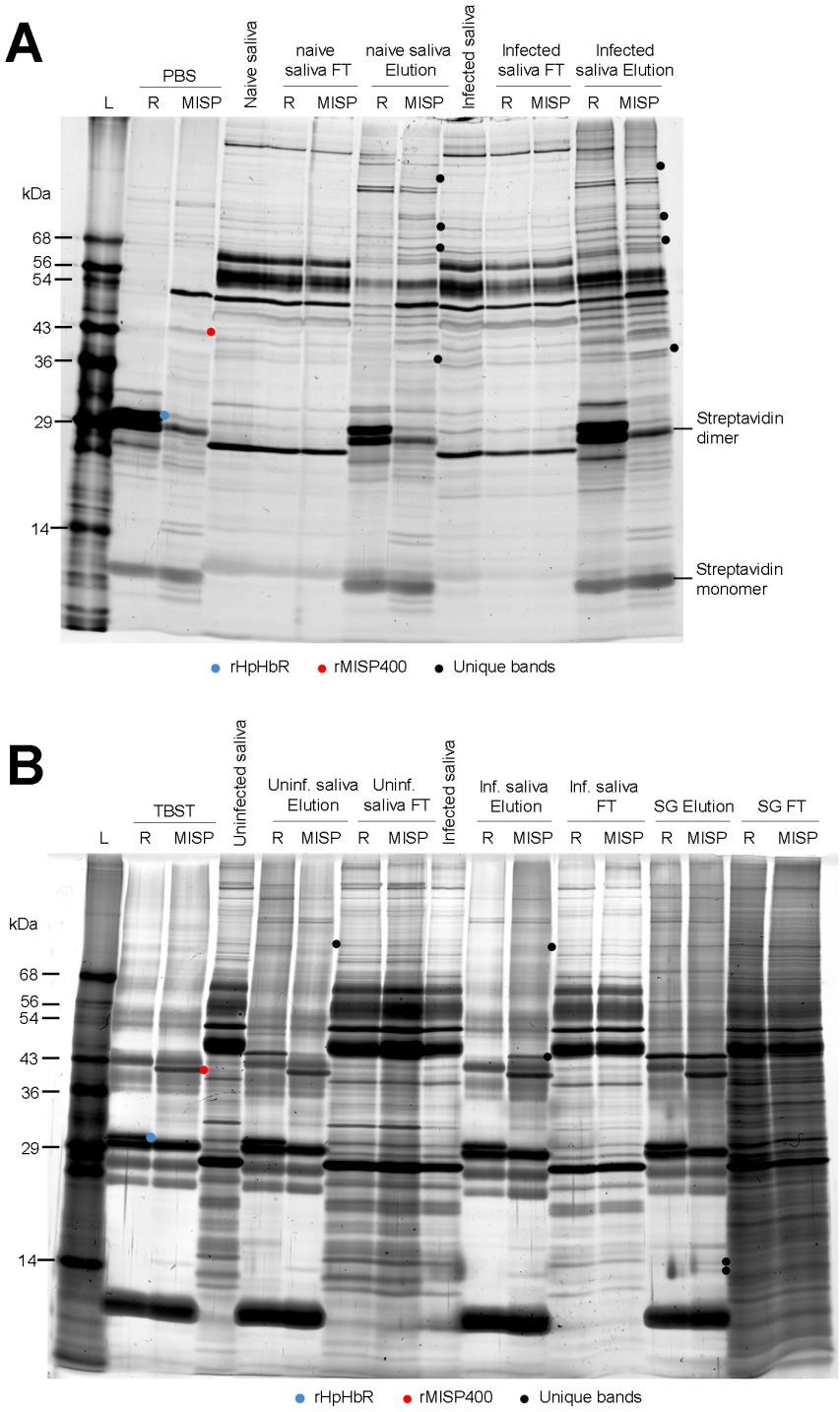


Figure 3.35. Pulldowns assays on MISP400. Silver-stained SDS-PAGE gels with pulled-down samples from two replicates **(A, B).** Biotinylated recombinant HpHbR (R) and MIS400 (MISP) incubated with PBS, uninfected tsetse saliva (uninf.), infected saliva (inf.) and whole infected tsetse salivary glands (SG). Direct samples, flow-through (FT) and elution. Dots are placed at the right of the band to point out: rHbHpR (blue), rMISP400 (red), unique pulled-down bands (black).

3.3 Discussion

3.3.1 *T. brucei*-infected tsetse saliva is enriched with trypanosome GPI-anchored surface glycoproteins.

Transmission of vector-borne pathogens usually requires insect saliva as a "vehicle", but it is also accompanied by a series of soluble components from both the parasite and the insect. For instance, *Leishmania* spp. secrete within sand flies abundant promastigote secretory gel and exosomes, which are important virulence factors for the transmission and establishment of the parasite infection in the vertebrate host (Atayde et al., 2015; Giraud et al., 2018; Rogers, 2012). Here, a semi-quantitative proteomics approach was used to analyse the composition of *T. brucei*-infected tsetse saliva with the aim of identifying potential soluble factors involved in parasite transmission. Besides hits from *Glossina* and *T. brucei*, proteins from the bacterial symbiont, *S. glossinidius* (Cheng and Aksoy, 1999), were also detected. Compared to naïve flies, *Sodalis* proteins seem to be more abundant in saliva during a trypanosome infection, as shown by western blotting (**Figure 3.6**). This could be explained in part by a higher cell permeabilisation led by the trypanosome infection, as *Sodalis* is normally found within epithelial cells (Cheng and Aksoy, 1999; Wang et al., 2013).

The proteomics analysis identified 27 different *T. brucei* proteins in infected saliva, of which 62.9% are likely cytosolic and cytoskeletal. The reasons behind the presence of trypanosome proteins in saliva are unknown; they could simply originate from dying parasites or their release could be controlled by specific mechanisms. Interestingly, 7 of these proteins (41.2% of intracellular hits) have been described as part of the *T. brucei* exosome (Szempruch et al., 2016), but I did not investigate whether trypanosomes release exosomes during development in the tsetse SG. Notably, *T. brucei*-infected saliva is particularly enriched in trypanosome GPI-anchored surface proteins (37.1%); i.e. BARPs (Tb927.9.15520, Tb927.9.15530, and Tb927.9.15570, similar to a previous report (Kariithi et al., 2016), several VSGs (discussed below) and MISP, which this thesis describes as the first family of invariant surface proteins in *T. brucei* MCFs.

In summary, the *T. brucei*-infected infected tsetse saliva is not only composed of infective MCFs (and EMFs in a low proportion), but it is rather a cocktail of *G.*

morsitans salivary, *S. glossinidius* and *T. brucei* proteins. How this complex composition modulates parasite transmission has yet to be investigated.

3.3.2 Do metacyclic VSGs recombine?

The proteomics analysis identified peptides from several VSG species, which are considered as expressed only in the blood stages. Although the parasite re-expresses VSGs only in the metacyclic stage (Tetley et al., 1987), their role during development within the tsetse or during transmission remains unclear. The identification of mVSG peptides in infected saliva is not surprising given that infected SGs contain thousands of quiescent metacyclic cells, which may release VSGs into the SG lumen or die before they get transmitted. The release of mVSG may occur partly by the action of the parasite GPI-PLC (Grandgenett et al., 2007; Gruszynski et al., 2003), which was found to be expressed in MCF in previous studies (Christiano et al., 2017; Rotureau et al., 2012). Interestingly, among the VSG species detected, only one corresponds to a canonical mVSG (mVAT4), which has not been previously found as protein, but its expression site has been well characterized by others (Pedram and Donelson, 1999). Only a similar study detected mVAT5 in tsetse saliva infected with a different *T. brucei* strain (Kariithi et al., 2016). Although a few studies reported BSF expressing mVSGs from BES at low levels (Alarcon et al., 1994), no study has described the expression of BSF VSG proteins in MCFs. This supports the assumption that, though recombination may occur from MES to BES, it does not appear to occur in the opposite direction since MES lack the 70-bp repeats upstream the mVSG gene (Alarcon et al., 1994; Pedram and Donelson, 1999). Surprisingly, however, we identified at least 4 VSG species that have been reported as expressed in BSFs, including MiTat 1.2 VSG 221 (identified by 3 unique peptides) (Cross et al., 2014). This suggests that either 1) MCFs could activate BES in low levels (unlikely since MCFs do not appear to express ESAGs strongly associated with BSFs (i.e. ESAG6/7) (Christiano et al., 2017)), or 2) there is recombination at the MES to replace the *mvsg* with BSF *vsg*, or 3) there is recombination between *mvsg* and BSF *vsg* genes leading to the formation of mosaics. In addition, we cannot rule out that the undetermined VSG repertoire of the TSW196 strain (Paindavoine et al., 1986) (used in this study) could be different to that in other strains (e.g. Lister 427;(Cross et al., 2014)) and lead to a misinterpretation of these results. Importantly, previous studies have found similar results although they have been overlooked (Christiano et al.,

2017; Kariithi et al., 2016; Savage et al., 2016; Telleria et al., 2014) (summarised in **Table 3.3**). For example, at least 2 BSF VSGs were identified in a *T. b. brucei* EATRO 1125-infected saliva proteomics (Kariithi et al., 2016); one BSF *vsg* was found to be highly transcribed in a *T. b. brucei* RUMP503 SG infection (Telleria et al., 2014); and 15 BSF *vsg*s had similar transcription levels than *mvsg*s (and a few were translated into protein) in *T. b. brucei* Lister 427 MCF overexpressing *rbp6* (Christiano et al., 2017).

3.3.3 The Metacyclic Invariant Surface Proteins (MISP)

We characterised the MISP family studying, among others, gene and protein expression during the parasite life cycle, cellular localisation, essentiality for parasite development in the tsetse and determined the crystal structure of the MISP360 homolog. The majority of these studies were conducted alongside with BARP as a known family of surface proteins expressed by EMFs in SGs (Urwyler et al., 2007). MISPs were previously included in the trypanosome surface phylome (Jackson et al., 2013), within the clade 'iv' of the Fam50 family, which also includes BARP, the *T. congolense* GARP (Bayne et al., 1993; Beecroft et al., 1993) and CESP (Sakurai et al., 2008). With the exception of *T. vivax*, most of the disease-relevant species of African trypanosomes such as *T. brucei* (including *T. b. gambiense* and *T. b. rhodesiense*), *T. congolense* and *T. evansi* encode *misp* genes. The conservation of MISPs across the family suggests an important role in the parasite's life cycle within the tsetse or during transmission and may reflect a conserved function.

Importantly, phylogenetic analyses have shown that MISP have evolved in two subfamilies (MISP-A and MISP-B) which, despite sharing high sequence conservation, diverge in many of the predicted characteristics. First, we have shown that they diverge in a few domains which mostly localise around the molecular pocket on the apical face of the protein, suggested to be involved in molecular recognition. This would imply minor differences in function between subfamilies. Second, MISP-B members vary in *N*-glycosylation sites having either 2 or none, compared to the conserved 1 site in MISP-A. In addition, although both subfamilies have a strong prediction for signal peptide, MISP-B seems to have a very distinct sequence that may be potentially cleaved at a different site compared to that of MISP-A. Lastly, a

distinct characteristic between subfamilies is the presence of C-terminal repeats of 26 residues found in MISP-A and absent in MISP-B, which have been shown here to be important for the display and exposure of MISP above the mVSG coat. Despite minor differences between subfamilies, however, there are also many conserved features predicted such as GPI-anchor peptides, protease cleavage sites, and overall 3D structures of the N-termini.

3.3.4 Production of MISP recombinant proteins for the study of MISP

To study MISP, it was fundamental to raise an antibody specific against the proteins, so it could be used in immunodetection. The approach of choice was immunisation of a rabbit with a complete MISP recombinant protein. It was hypothesised that, due to the high sequence conservation between isoforms, the polyclonal antibody would cross-react with them all. For technical reasons, the isoform of choice was MISP380 as its expression levels turned out to be higher than the other MISP isoforms. This was recombinantly produced in *E. coli*, purified and used for immunisation. Immune anti-serum was further processed to isolate IgGs with specific binding with MISP and depleted from *E. coli* interactions. The polyclonal anti-MISP380, broadly used in this thesis, proved to be specific for MISP and to cross-react with, at least, the isoforms MISP360 and MISP400. The antibody did not react with *E. coli* proteins or *T. brucei* PCF lysates as expected.

The use of recombinant proteins was not only useful for the production of a polyclonal antibody, but also for structural studies. The crystal structure of MISP360 N-terminus was solved from crystals of a recombinant protein produced in *E. coli.* In addition, the production of trimmed versions of MISP360 (N-terminus, C-terminus and complete protein) will allow future NMR studies to determine the structure of MISP C-terminus and the motion dynamics between the N- and C- terminal domains, even within the context of a plasma membrane (by using nanodiscs, (Munoz-Garcia et al., 2018)).

3.3.5. Expression studies on MISP and BARP

We have confirmed that *misp* and *barp* gene expression is upregulated in *T. b. brucei* SG stages as previously described (Savage et al., 2012; Savage et al., 2016; Telleria et al., 2014). However, semi-quantitative RT-PCR analysis has allowed, for the first time, a detailed study of the individual expression of *misp* homologs by exploiting the different number of repetitive motives at the C-termini. It was found that the upregulation of *misp* is not homogeneous across homologs but biased towards a preferential expression for *misp-B* paralogs (the shortest *misp400* and *misp420*), at least at the transcript level. With the exception of *misp440* that seems to have a low but stable expression, *misp-A* genes are only transcribed during the late stages of development in the SGs. It is unknown how the expression of these genes is regulated despite being part of a gene array that is likely polycistronically-transcribed and with >99.7% conservation in both UTRs and intergenic sequences. Differences in homolog transcript levels, however, may not reflect changes in protein isoform abundances, which remain unknown as the anti-MISP polyclonal antibody is expected to cross-react with all isoforms due to high sequence conservation among them. In fact, this polyclonal antibody can detect, at least, MISP360 and MISP380 by western blotting, and MISP360 and MISP400 (not shown) by cell immunostaining. Interestingly, although PV-dwelling trypanosomes produce large amounts of *misp* and *barp* transcripts (this work and Urwyler et al., 2007), the proteins are only translated in SG forms and localise evenly over the cell surface, as shown by immunostaining of individual SG forms. Instead, immunostaining shows that all PV trypanosomes express EP-procyclins, even in EMFs, which is a surprising finding (**Figure 3.33**) (Sharma et al., 2008). While BARP protein expression peaks in EMFs and is lost during metacyclogenesis, MISP proteins start being expressed in EMFs but reach their highest expression in MCF, which suggests MISP is a major surface protein expressed by MCF. However, the copy numbers of MISP in relation to VSG remains to be determined. Strikingly, anti-MISP staining also occurs in live MCF cells (free of potential fixation artefacts), suggesting that MISP display immunogenic epitopes above the mVSG protective coat. Interestingly, despite the previous observation that BARP is absent in MCFs (Rotureau et al., 2012; Sharma et al., 2008), we detected low levels of BARP in MCFs with variable localisation, a phenomenon that could be explained by differences in parasite strain and/or infection timing. While a minority of cells showed a surface localisation of BARP, most had re-distributed the protein to the flagellum. An alternative explanation is that most metacyclic cells express low

(undetectable) amounts of BARP on the parasite surface, while its high concentration in the flagellum makes it easier to detect by fluorescent microscopy. Nevertheless, a similar flagellar localisation was observed in BSF (Jackson et al., 2013) and PCF (**Figure 3.24**) expressing ectopic MISPs, although, in the latter, it regained surface location once the cells differentiated into MCFs within the tsetse.

3.3.6 Insights into the protein structures of MISP and BARP

In order to gain more insight in to the function of MISP, the crystal structure of the N-terminal domain of the MISP-360 isoform (Tb427.07.360) was solved. It revealed to have a triple helical bundle structure resembling that of other trypanosome surface proteins such as the *Tb/Tc*HpHbR (Higgins et al., 2013; Lane-Serff et al., 2014), *Tc*GARP (Loveless et al., 2011) or a VSG monomer (Bartossek et al., 2017; Freymann et al., 1990). The mature MISP, however, contains an additional, 127 residue C-terminal tail predicted to be highly disordered. In the absence of definitive structural data, I modelled the C-terminus, which indicated that it might serve as a long tether effectively elevating the N-terminal head group from the surface of the parasite membrane. This explains in part the recognition of MISP by antibodies shown in live immunostainings. It is tempting to speculate that this extension facilitates a biologically relevant interaction between the parasite and its environment such as promoting the acquisition of nutrients, serving as an adhesion or decoy, or modulating a response in the host. The previously characterised *Tb*HpHbR has been shown to lie partially within the VSG layer of the BSF and allows the trypanosome to acquire nutrients from the blood of the mammalian host. The mechanism by which it acquires nutrients is facilitated by a C-terminal extension that increases its height making the ligand-binding site accessible by protruding above the VSG layer (Higgins et al., 2013). Since MCF express mVSGs (Tetley et al., 1987) along with MISP, its C-terminal extension could support a similar role though the substantial length of the extension would likely render cellular uptake of a nutrient payload difficult. However, the unstructured extension is likely ideally suited to support an adhesion role for MISP, similar to the function described for bacterial adhesins like the *Haemophilus* Cha proteins (McCann et al., 2014). Interestingly, the only study showing evidence of a trypanosome surface protein with an adhesin function is on *Tc*CESP, which is also part of the Fam50 family of trypanosome surface proteins (Sakurai et al., 2008). The

structural analyses of MISP360 revealed the existence of a predominantly hydrophobic pocket in the apical end. It is possible that the pocket may allow MISP360 to coordinate a partner molecule, consistent with the putative role as an adhesin. Despite the high sequence identity, the residues forming and surrounding the pocket lacked sequence conservation and resulted in a distorted cavity between the homolog models that may reflect the capacity to coordinate a structurally diverse set of ligands. Furthermore, the homologs differ in the length of the C-terminal region, originated from the variable number of 26-residues motifs. This region could offer the parasite a robust mechanism to engage a variety of cell types or provide a sequential binding effect to enable tighter adhesion. It is noteworthy that these putative unstructured C-terminal regions contain lysines and arginines indicating an inherent susceptibility to proteolysis consistent with predictions using the PROSPER software (Song et al., 2012). This putative proteolytic susceptibility may be exploited by the parasite to release the molecular interactions anchoring it to the epithelial cells of the salivary glands of the tsetse and thereby facilitating transmission.

3.3.7 The metacyclic surface glycocalyx

It has been assumed that the MCF surface is structurally and functionally equivalent to that of BSFs as they show the same thick VSG (electron-dense) layer in TEM sections (Kolev et al., 2012; Ooi et al., 2018). A few BSF VSG structures have been determined by X-ray crystallography (Bartossek et al., 2017; Blum et al., 1993; Freymann et al., 1990), but no mVSG structures have been reported so far. Based on protein sequence, however, it is likely that mVSGs should fold in a similar way and form the same homodimer structures. This implies mVSGs also form a dense macromolecular barrier that protects the parasite and masks most invariant surface proteins from the host's immune system. Most of the characterised BSF invariant surface proteins are attached via transmembrane domains and localize in the flagellar pocket, with the exception of the GPI-anchored transferrin and haptoglobin-haemoglobin receptors (Lane-Serff et al., 2014; Tiengwe et al., 2017). However, our crystallographic and *in vivo* labelling evidence shows that MISP appear to be displayed above the mVSG coat, suggesting that, unlike the blood stages, MCFs allow the homogeneous expression of other GPI-anchored invariant proteins on the surface. Thus, I hypothesise that MISP could be targeted by the host's immune system, although it remains to be investigated if the brief exposure of MISP in skin

(Caljon et al., 2016), right after transmission by an infected fly, is able to trigger a protective and sustained immune response.

3.3.8 Function of MISP

MCF are known to be cells with arrested cell cycle and low transcription activity and therefore, detecting high levels of *misp* and *barp* transcripts suggests that these proteins may play an important role in parasite development in the SG, or during the early stages of infection of the mammalian host. To gain insight into the possible role(s) of MISP, I partially depleted *misp* expression in a mutant cell line. A clonal mutant procyclic line with induced knockdown did not experience a *misp* down-regulation and did not suffer growth defects. However, when these cells progressed through the life cycle in the tsetse, *misp* levels were significantly down-regulated in PV and SG parasites, as also shown by indirect immunostaining (**Figure 3.34**). The reduction of MISP, however, did not seem to have an effect in parasite viability or infectivity in the tsetse, suggesting that MISPs may not be essential in development and perhaps play a role later in transmission or early infection of the mammalian host when MCF are found in the skin (Caljon et al., 2016). However, it cannot be ruled out that MISP knockdown was insufficient or that it has a limited role in tsetse infectivity. Remarkably, this is the first successful knockdown carried out in *T. brucei* SG forms.

A conditional-null mutant for the 5 *misp* genes would be the ultimate tool to study MISP function. However, the complete knockout is technically unfeasible using the classic genetic approaches, since the complete removal of the *misp* array would also remove the unrelated genes in between. These could be potentially added back (along with *misp*) but achieving a proper regulation of their expression would be challenging too. Fortunately, the recently arisen CRIPSR/Cas9 gene editing technology will likely allow the complete knockout of *misps* without affecting the expression of unrelated genes in the same array. Although the technology has been optimised for several systems such as *T. cruzi* and *Leishmania* (Peng et al., 2014; Sollelis et al., 2015) in the last years, it has been more challenging to implement in *T. brucei* (Beneke et al., 2017; Rico et al., 2018; Vasquez et al., 2018).

I also investigated putative MISP binding partners. For technical reasons, a recombinant biotinylated MISP400 was used as a model. Biacore resonance studies showed that there seems to be no binding of MISP400 to any component from the two main environments metacyclic cells can be in: tsetse saliva and mammalian serum. However, pulldown assays suggested that MISP400 may interact with a few protein components from tsetse saliva (both uninfected and *T. brucei*-infected). These unique bands shown in the pulldown SDS-PAGE have been isolated and are currently under analysis by mass spectrometry. The identification and further validation of these pulled-down proteins may be key in the path towards the identification of the function of MISPs.

Up to now, the function(s) of MISP remains unknown and can only be hypothesised. Due to the uniqueness of the protein sequences, no functions or domains can be found by sequence homology. The solved crystal structure showed that, despite MISP are structurally similar to other surface trypanosome proteins (e.g. BARP and GARP, both with unknown functions), the lack of conserved residues and low identity conservation suggests that these proteins may not have similar biological roles. From the MCF surface model generated, it can only be hypothesised that MISP may have a potential role in adhesion, ligand recognition or decoy for the host's immune system.

3.3.9 Uses of MISP and future work

MISP and BARP have the potential to be used as molecular markers for metacyclogenesis. There is the need for the development of molecular tools to identify transmissible tsetse (i.e. carrying MCFs in SGs) in disease-endemic areas. Only a limited number (~0.1%) of flies trapped in the field carry infectious parasites (Auty et al., 2012; Jamonneau et al., 2004). Unlike current technologies based on manual and rather laborious dissection or PCR-based methods from pools of flies (Ngomtcho et al., 2017), exploiting immune-detection of MISP proteins could help to identify disease transmission areas requiring vector control intervention(s). Thus, immunodetection of MISP and BARP proteins would greatly facilitate identification of disease-transmissible tsetse flies.

On the other hand, based on the results of this thesis, MISP can be proposed to be a candidate for the potential development of a transmission-blocking vaccine against HAT or AAT produced by either *T. brucei brucei*, *T. congolense* or *T. evansi*. As shown in this chapter, MISP are the first invariable proteins reported to be expressed on the surface of MCF, immunogenic and exposed to antibodies. Immunisations with MISP could potentially provide the mammalian host with antibodies capable of neutralising tsetse-injected MCF before they differentiate into BSF and start replicating. Although HAT is heading towards elimination in 5-10 years and a vaccine against sleeping sickness is probably no longer needed, this proof-of-concept vaccine would open the door for these methodologies to explore metacyclic invariant surface proteins as vaccine candidates against animal trypanosomiasis.

CHAPTER 4

Generation of *T. brucei* metacyclic forms

in vitro for the study of MISP

4.1 Introduction

In Chapter 3, MISPs were presented as a novel family of invariant surface proteins expressed by the metacyclic trypomastigotes of *T. b. brucei*. Although many aspects of MISP have been characterised, more studies are required to understand the biological function(s) of these proteins. Most of these studies would require large amounts of MCF cells, which are difficult to obtain from infected tsetse. Thus, there is the need to explore alternative ways to obtain MCF cells. In 2012, a serendipitous discovery occurred while studying *T. brucei* RNA-binding proteins. It was found that the overexpression of the RNA-binding protein 6 (RBP6) in PCF triggered cells to differentiate into EMFs and MCFs, fact which does not normally occur *in vitro* (Kolev et al., 2012). RBP6 overexpressor cells experienced a progressive decrease in the proportion of PCF because some became EMFs as short as 2 days post-induction of the transgene, resembling those parasite stages found in the tsetse PV, and a peak of MCFs at 10 days post induction. MCFs were validated based on morphology, relative positions of nuclei and kDNA, mitochondria morphology, endocytosis rate and visualisation of the VSG coat by TEM, among others. In addition, MCF infectivity was proven when purified MCF were injected into mice, leading to high levels of parasitaemia in infected animals. Remarkably, since the Kolev et al. 2012 publication, no other works have been reported by other groups using this system to generate *T. brucei* EMFs and MCFs *in vitro*, presumably due to difficulties in reproducing the same protocol by different groups.

In this chapter, I first describe the reproduction and characterisation of this system to generate *T. brucei* MCF *in vitro,* with special focus on the expression of developmentally regulated major surface glycoproteins, including MISP and BARP. Remarkably, the generation of a BSF-like form (not reported by Kolev et al. 2012) was also observed, presumably originated from MCFs. I also show the purification of MCF by ion-exchange chromatography from large volumes of culture, which could be further used in downstream applications. Nothing is known about the biological role of RBP6 in *T. brucei* development in the tsetse. Several published and unpublished (this thesis) transcriptomics works (Christiano et al., 2017; Savage et al., 2016) have described that *rbp6* is up-regulated in parasites residing in PV and SG compared with MG and blood stages, suggesting that it is important for completion of *T. brucei* developmental cycle in the fly. In this chapter I studied the role of RBP6 *in vivo* and also identified other *T. brucei* genes (besides *rbp6*) important for the establishment of

a SG infection and differentiation into MCF. These include several folate transporters, a glutamate dehydrogenase and a hypothetical protein of unknown function.

4.2 Results

4.2.1 Creation of novel *T. brucei* cell lines overexpressing RBP6

To reproduce the system for the generation of MCFs *in vitro*, several mutant *T. brucei* cell lines overexpressing *rbp6* were created. Since it is unknown if the expression level of the ectopic *rbp6* has an influence on the differentiation process, two plasmids with different promoters driving the expression of the transgene (tetracycline inducible) were used to ensure variable expression levels of the transgene: pDEX577 (Kelly et al., 2007) using a T7 promoter for medium-high expression, and pMS (derivate of pLEW100v5, (Serricchio et al., 2013) using a rDNA promoter for high expression (**Figure 4.1**). The ectopic *rbp6* gene was either tagged with an HA epitope to allow immunodetection or with 3 consecutive HA epitopes followed by a superfolder GFP (sfGFP) to facilitate *in vivo* imaging as well as immunodetection. Both tags were fused to the RBP6 N-terminus (upstream the *rbp6* gene) as it was shown to be less involved in the overall protein structure in a model generated using I-TASSER (Roy et al., 2010; Yang et al., 2015; Zhang, 2008), compared with the C-terminus (**Figure 4.1**). No signal peptides were predicted in RBP6. In addition, I created RBP6 overexpressors in two different *T. brucei* PCF strains, Lister 427 29:13 (non-fly-transmissible; used in (Kolev et al., 2012) and AnTat 1.1 90:13 (fly transmissible, (MacGregor et al., 2013); used for the characterisation of MISP in Chapter 3. After transfection and drug selection, clonal transfectant cell lines were screened and selected for homogeneous and high expression of ectopic RBP6. Cells were shown to express ectopic RBP6 upon Dox induction using a polyclonal anti-HA antibody, which showed it localised in the cytoplasm in small and diffuse microgranules (**Figure 4.2**).

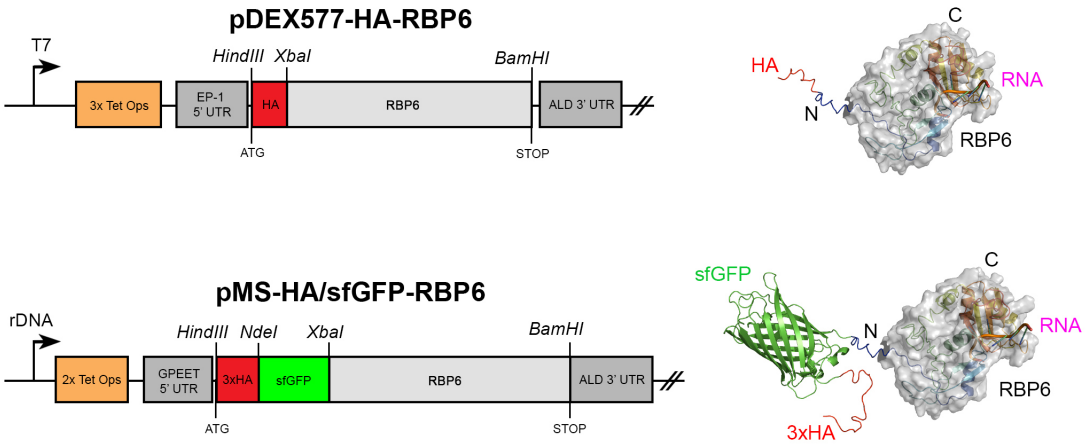


Figure 4.1 Overexpression of ectopic RBP6 in *T. brucei* PCF. DNA constructs (left) used for the overexpression of RBP6 fused to an HA epitope at N-terminus (top) or 3xHA epitopes and superfolder GFP. Protein cartoon models (right) of the expected 3D structure of tagged RBP6 (molecular surface in grey) bound to a putative RNA molecule (to predicted binding site). Domains highlighted: HA epitopes (HA, 3xHA), superfolder GFP (sfGFP), N- and C- termini of RBP6 (N, C).

Alternatively, cells expressing RBP6 fused to sfGFP were selected by imaging the fluorescent protein, which co-localised with the HA label (**Figure 4.2**). This goes in agreement with the localisation results in TrypTag (www.tryptag.org, (Dean et al., 2017) (high-throughput endogenous mNeonGreen fluorescent protein N-tag). Fused tags may sometimes make proteins mislocalise by affecting folding or transportation, so a correct localisation is an essential requirement to ensure functionality. Importantly, induced PCF cells from the 3 generated cell lines showed differentiation into EMF and MCFs, proving the ectopic tagged RBP6 to be functional (**Figure 4.2**). It is worth mentioning that EMF cells generated *in vitro* are equivalent to those found in infected tsetse and described in detail in Chapter 3, as they are mainly identified for having the kDNA in the anterior end of the parasite (**Figure 3.22** and **4.2**). However, because the differentiation rate *in vitro* (i.e. PCF-EMF-MCF) occurs too quickly, biochemical and cellular differences may exist between these *in vitro* forms and the ones found in infected flies.

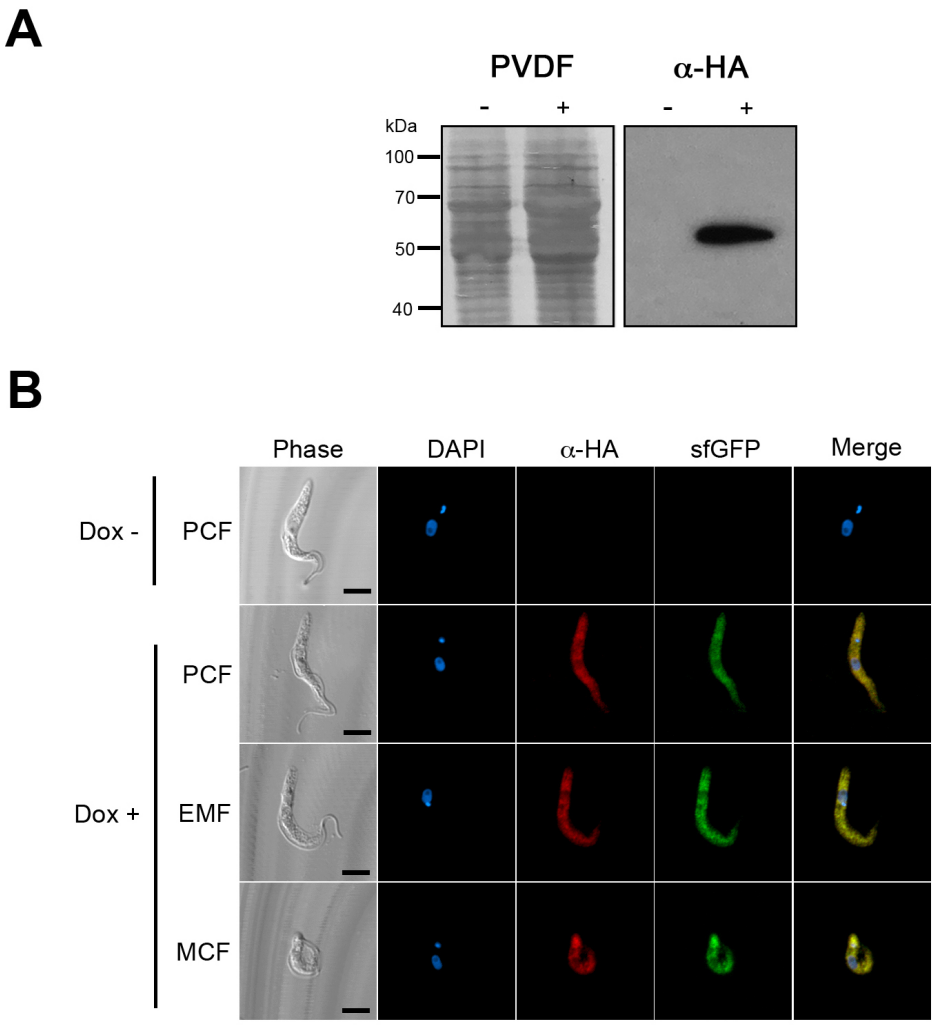


Figure 4.2. Cellular localisation of ectopic RBP6. (A) Western blotting of RBP6 overexpressor cells either uninduced (Dox -) or induced (Dox +). **(B)** Immunostaining of previous cells stained with DAPI (blue), probed with anti-HA antibodies (red), sfGFP in green and merged image. Representative images of cultured procyclics (PCF), *in vitro*-generated epimastigotes (EMF) and metacyclics (MCF) after 3 days post-induction. Scale bars= 5 μm.

4.2.2 Differentiation dynamics of RBP6 overexpressor cells

A time-course experiment was set to determine the timing of differentiation of PCF to EMF and MCF. Since induction, samples were taken each day to estimate the proportion of PCF-EMF-MCF and the cell cycle phase of PCF and EMF cells in culture (**Figure 4.3-A**), in relation to changes in growth rate (**Figure 4.3-B-C-D**). After 24

hours post-induction, ~23% of PCF (characterised by a “torpedo”-shaped body and a kDNA posterior to the nucleus and far from the posterior end) triggered differentiation into EMF since they started migrating the kDNA to an anterior position relative to the nucleus. This proportion of EMF remained more or less stable during the time course analyses. Also, no significant changes in the proportion of dividing PCF and EMF cells were detected (**Figure 4.3-A**). As soon as 3 dpi, the first P-MCF and MCF cells, short with a rounded posterior end and with the kDNA at the tip of the posterior end, appeared in culture in a proportion of ~37%. MCF are known to be quiescent cells that cannot divide but differentiate into BSF or die within a few days. This may explain why in the days following the peak of MCF production, the culture becomes ~90% PCF again as the dividing PCF overcome the terminally-differentiated MCF (as reported in (Kolev et al., 2012). However, it was also observed that some MCF were significantly longer than usual (**Figure 4.4** and **4.5**) and some were even in the process of division; i.e. having 2 nuclei and kDNA. Thus, I considered these as MCF that potentially differentiated into BSF (BSF-like) as a result of the RBP6 overexpression or as an adaptation to the culture conditions. Although difficult to prove, these BSF-like could have differentiated into PCF and thus account for the re-emergence of PCF towards the end of experiment (**Figure 4.3-A**). In parallel, cell growth was recorded at each day post RBP6 induction. Compared to uninduced control cells, induced cells experienced a severe decrease in growth rates since induction started (**Figure 4.3-B-C**). Cell differentiation seems to be the most likely cause, as cells do not divide while undergoing this process. Despite the overall reduction, big changes in the growth rate can be matched with the differentiation event peaks (**Figure 4.3-D**). The slowest cell growth rates were found at day 0-1 and day 3-4 post-induction, corresponding with the main differentiation to EMFs (1 dpi) and MCF (3 dpi) events.

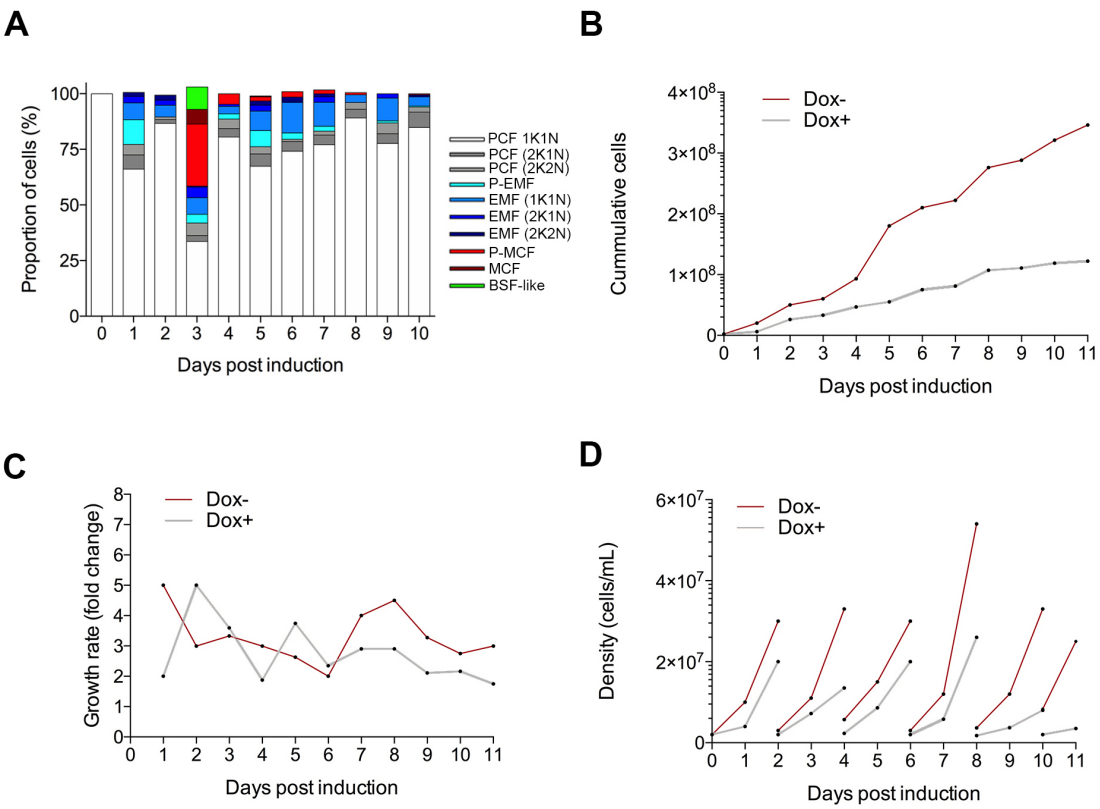


Figure 4.3 Overexpression of RBP6 triggers the differentiation of *T. brucei* cells *in vitro*. **(A)** Distribution of parasite stages during a time course (10 days) differentiation process of RBP6 overexpressor PCF cells (3xHA-sfGFP-RBP6 in strain Lister 427 29:13). ~99% control cells remained in the procyclic stage. Key: PCF (procyclic), EMF (epimastigote), MCF (metacyclic), BSF-like (bloodstream form like); KN refers to the number of kDNA and nuclei. The 'pre-' (P-) forms are intermediate stages characterised by the partial rotation of the kDNA and similar cell size than the preceding stage. **(B)** Growth curve of induced (Dox +) RBP6 overexpressor cells compared to uninduced (Dox -). Cultures split by 1/10 dilution at day 2, 4, 6, 8 and 10 post-induction. **(C)** Cumulative cell growth and **(D)** culture growth rate (fold change).

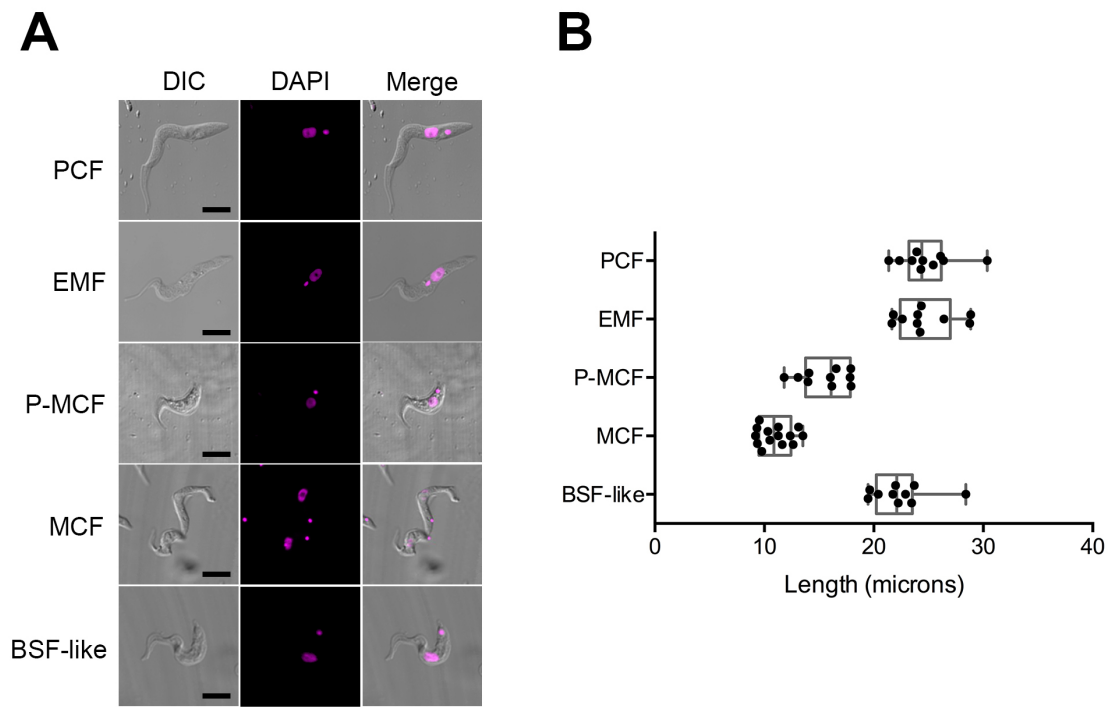


Figure 4.4. *T. brucei* life stages generated *in vitro* during RBP6 overexpression. **(A)** Representative images of procyclic cultured forms (PCF), pre-epimastigotes (P-EMF), epimastigotes (EMF), pre-metacyclics (P-MCF), metacyclics (MCF) and bloodstream form-like (BSF-like) stages. **(B)** Box plot of the measurements of total cell length of the previous stages. The boxes limit quartiles, the inner lines represent the medians and the whiskers represent variability. Scale bars = 5 μm.

4.2.3 Ion-exchange purification of *in vitro* MCF and BSF-like cells

The *in vitro*-generated MCF and BSF-like can be purified from the rest of parasite forms (i.e. PCFs and EMFs) by ion-exchange chromatography using a cellulose DE52 column based on the diethylaminoethyl (DEAE) tertiary amine functional group (Fox et al., 1986). Although not well understood, separation appears to occur thanks to the different surface net charges of the cells, mainly provided by the surface glycoprotein coats. MCF and BSF-like (mainly coated by VSGs) passed through, while PCF and EMF got trapped in the column. Induced RBP6 cultures at 3 dpi (found to have the highest percentage of MCF and BSF-like, **Figure 4.3-A**) could be scaled up in volume and purified using this method, yielding an average of ~100

MCF cells/mL. Remarkably, this purification method was applied to control uninduced cells, which yielded a 100% retention of cells under the same conditions. This supports the strong reliability of the purification method and provides one more evidence of parasite differentiation in induced cells.

4.2.4 Developmental expression of major surface glycoproteins

Morphology and relative position of nuclei and kDNA are not fully reliable parameters to characterise cell differentiation, especially in cultured parasites forced to differentiate *in vitro*, as morphologies within a life stage may greatly vary. Thus, there is the need to use molecular markers to track differentiation events more precisely and with higher confidence. It is well understood that *T. brucei* expresses stage-specific major surface glycoproteins to adapt to the different environments it encounters during its life cycle. Briefly, as discussed in **Chapter 3**, BSF express VSGs for antigenic variation (Horn, 2014) which are then replaced by procyclins (GPEET and EP) when it differentiates to PF in the tsetse MG (Roditi et al., 1987). EP-procyclin expression is maintained in PV mesocyclics and EMFs, although these may start expressing BARP to prepare for migration to the SGs (Sharma et al., 2008). Attached EMFs in the SG are known to be free of EP but to express a BARP coat instead (Urwyler et al., 2007) and this thesis) alongside MISP (**Chapter 3**). P-MCFs switch on the expression of mVSGs which will fully coat the mature MCF cells (Rotureau et al., 2012; Tetley et al., 1987) along with MISP and some BARP remains (**Chapter 3**). Accordingly, the detection of these molecules in the RBP6-differentiated cells could be a valuable tool to better understand the differentiation dynamics, assuming that these *in vitro* cells follow the same expression patterns as in the tsetse. Induced cells at 3 dpi were searched for the molecular markers BARP, MISP and CRD (Cardoso de Almeida and Turner, 1983; Zamze et al., 1988) by both immunofluorescence using fixed cells and western blotting in cell lysates.

The blottings were only found positive for anti-MISP in induced cells compared to uninduced (**Figure 4.5-A**). However, the apparent molecular mass of the components recognised by the antibodies is ~100 kDa, which is higher than the expected molecular mass of MISP (<40 kDa) on SDS-PAGE. This is not surprising, since all the recombinant versions of a complete MISP have always shown several bands in western blottings, being the 100 kDa band one of the major ones besides

the 40 kDa and suggesting a possible physical association (**Figure 3.17**). As explained in section 3.2.3.2, it is possible that MISP form aggregates through the C-termini which are not totally denatured when heated in sample buffer. Both anti-BARP and anti-CRD failed to detect the proteins in cell lysates despite the recombinant proteins positive controls reacted as expected. It is likely that the amount of protein is below the levels of detection, since the proportion of cells that could express these proteins is significantly small compared to the amount of undifferentiated PCFs.

The second approach consisted in immunostaining of individual fixed cells, which overcomes the problem of having small amounts of expressor cells. Induced and uninduced cultures at 3 dpi were fixed and probed with anti-MISP, anti-BARP and anti-CRD, the latter in combination with anti-EP procyclin. Both anti-MISP and anti-BARP failed to reveal any positive signal in either culture. Since at least MISP was found by western blotting, it might be possible that these molecules or epitopes are not accessible to antibodies on the surface compared to “real” fly forms. The anti-CRD, however, did clearly react with a few induced cells only. These cells were found to have the same size, morphology and DNA relative position considered in section 4.2.2 for the identification of MCF and BSF-like stages. Remarkably, all positive cells for anti-CRD were found to be negative for anti-EP procyclin (as expected), further supporting the identification of these cells as MCF or BSF. Interestingly, an anti-CRD positive cell (3 dpi) was found to be undergoing cytokinesis (**Figure 4.5-C**), thus confirming the possible presence of BSF-like forms in induced cells.

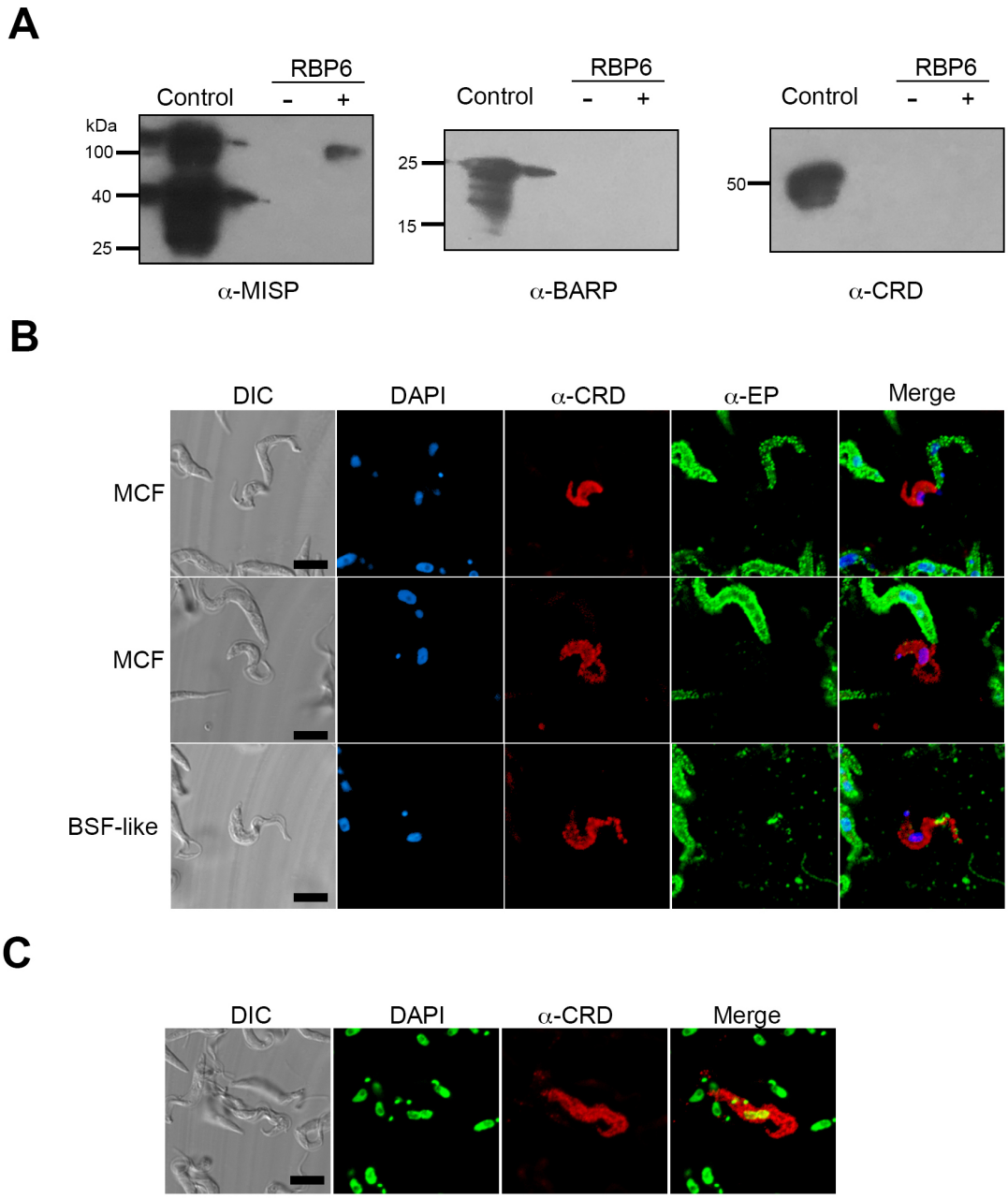


Figure 4.5. Detection of developmentally regulated surface glycoproteins in RBP6 overexpressor cells. (A) Western blotting of cell lysate equivalent to 10^7 PCF cells either uninduced (-) or induced (+), at 3 days post-induction, were probed with either anti-MISP, or anti-BARP or anti-CRD. Positive controls: recombinant MISP380 (anti-MISP), recombinant BARP (anti-BARP) and purified soluble VSG (anti-CRD). **(B)** Immunostaining of induced (3 dpi) RBP6 overexpressor cells with anti-CRD (red) and anti-EP procyclin (green). **(C)** Immunostaining of a dividing (2K2N) BSF-like cell from a 3dpi RBP6-induced culture, probed with anti-CRD (red) and stained with DAPI (green). Scale bars = 5 μm.

4.2.5 Developmental expression of RBP6

It has been shown that RBP6 is involved in differentiation from PCF to MCF *in vitro* upon overexpression, but its role in the biological context of the tsetse remains unknown. The RNA expression of *rbp6* was measured throughout the life cycle of *T. b. brucei* (strain AnTat 1.1 90:13, **Figure 4.6**) to identify the potential developmental regulation and to understand at which point its expression is naturally important to carry out its function. The RBP6 PCF overexpressor cells were used as control for detection, showing a ~2 fold up-regulation in induced cells compared to the uninduced ones. cBSF and MG procyclics from both 15 and 30 dpi, showed low levels of *rbp6* RNA which were comparable to that in PCF. *rbp6* expression seems to remain stable in PV stages at 15 dpi, but then it is significantly increased in PV parasites at 30 dpi (~3-fold). A high expression of *rbp6* seems to be maintained in SG parasites (~1.25-fold higher than PV at 30 dpi). These results suggest that, in the fly, RBP6 probably plays an important role in trypanosome differentiation at (and beyond) the PV, including metacyclogenesis inside infected SGs. This is reminiscent of the differentiation pattern observed *in vitro* during overexpression of RBP6.

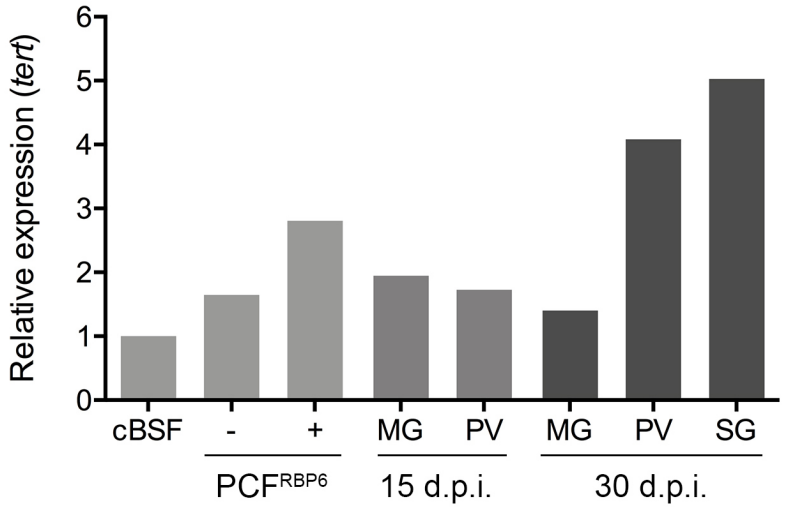


Figure 4.6. Developmental expression of RBP6 in *T. b. brucei.* RNA relative expression of *rbp6* normalised to the expression of *tert* during the life cycle of *T. b. brucei* strain AnTat 1.1 90:13 as cultured forms (BSF and PCF) or in the tsetse MG and PV at both 15 and 30 dpi and SG at 30 dpi.

4.2.6 *In vivo* phenotype of RBP6 overexpression in the tsetse

T. b. brucei Lister 427 29:13 RBP6 overexpressor cells (see Section 4.2.8.1 and 4.2.8.5), whose infectivity in the tsetse is limited and cannot progress beyond the PV (and thus cannot establish SG infections), were used to infect tsetse in two ways (**Figure 4.7-A**). First, flies were infected with uninduced cells and induction with Dox was started at 15 days post infection when the parasites have already colonised the PV (see section 4.2.8.1). We hypothesised that, if a defect in differentiation is what prevents Lister 427 from progressing beyond the PV, the overexpression of RBP6 at that point would induce differentiation of those PF and allow progression towards the SG. The second experiment consisted in pre-inducing the RBP6 overexpressors *in vitro* for 3 days (time point with more diversity of forms, **Figure 4.3**) and infect the tsetse with a mixture of PFs, EMFs and MCFs, expecting a potential early establishment of PV and SG infections bypassing MG colonisation. First, both experiments failed to generate SG infection rates due to the overexpression of RBP6 (**Figure 4.7-B**). MG infection rates did not greatly vary except for a slight increase in the proportion of heavy MG infections in experiment 1, probably due to the initiation of RBP6-induced differentiation into EMFs in the tsetse MG within the first days of infection. In experiment 2, however, PV infections were slightly increased as a consequence of inducing differentiation at 15 d.p.i. when the parasites had already migrated to the PV.

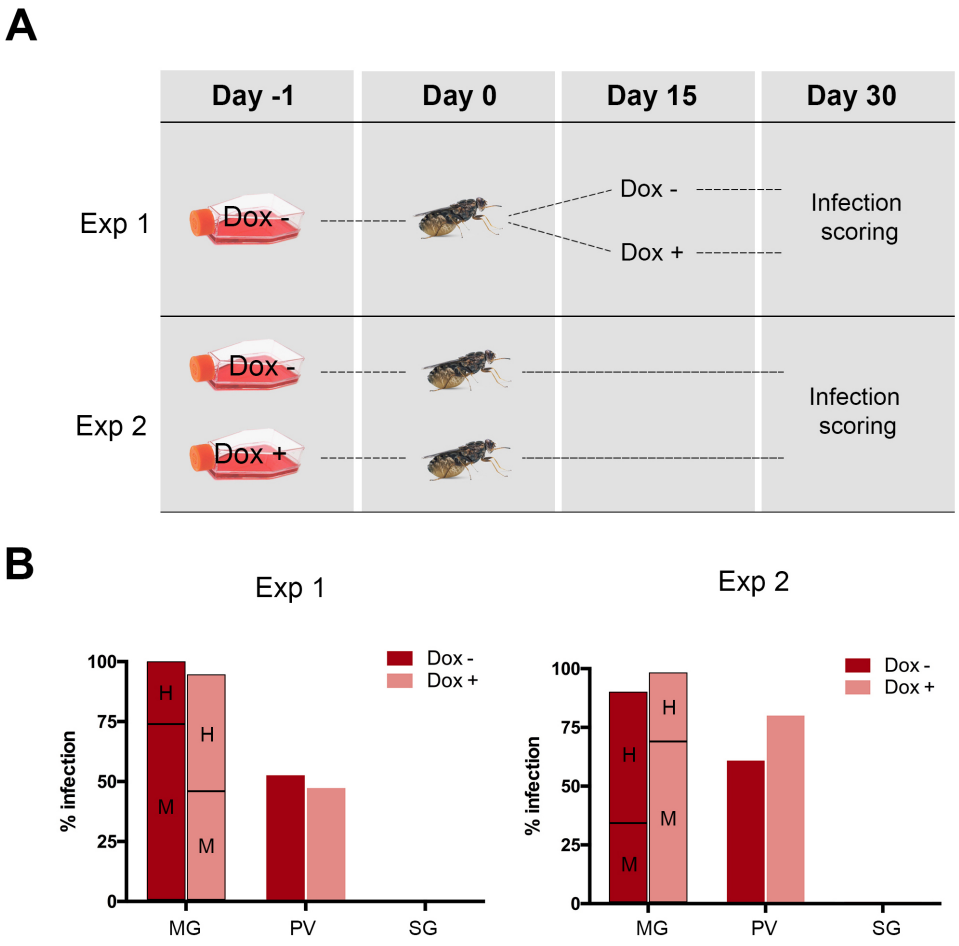


Figure 4.7. *In vivo* tsetse experiments with RBP6 overexpressors. (A) Experimental design for experiments 1 and 2. **(B)** Infectivity (percentage) of the RBP6 overexpressors (strain Lister 427 29:13) either induced (Dox +) or uninduced (Dox-) in the tsetse midgut (MG), proventriculus (PV) and salivary glands (SG) at 30 dpi. Midgut infection level distribution is noted as heavy (H) or medium (M).

4.2.7 Assessing the essentiality of RBP6 for parasite development in the tsetse

As shown in this thesis and by others (Kolev et al., 2012), RBP6 appears to be an important RBP for trypanosome differentiation during its transit between the fly's PV and the SGs. To further study the essentiality of RBP6 for *T. brucei* development in the tsetse, a *T. brucei* mutant cell line (fly-transmissible strain AnTat 1.1 90:13) was created to knockdown *rbp6* via RNAi. A unique fragment of the *rbp6* coding sequence was cloned into the vector pALC14 (MacGregor et al., 2013) for the expression of a stem loop short-hairpin dsRNA to induce gene knockdown under

tetracycline control (**Figure 4.8-A**). Knockdown levels were determined in PCF by qRT-PCR, revealing a ~50% transcript reduction as the strongest downregulation found in several cell lines screened (**Figure 4.8B**). Induced PCF cells did not experience any morphology or growth defects compared to uninduced cells *in vitro* (**Figure 4.8-C**). Tsetse flies were then infected and dissected at 30 dpi to determine infection rates in MG, PV and SG (**Figure 4.8-D**). Compared to uninduced cells, the downregulation of *rbp6* did not show a visible infection phenotype in the tsetse. Both groups of parasites were able to infect the MG (~80%) at comparable rates, and the PV at similar (or even slightly higher) rates than the control (~70%). Furthermore, no SG infections were observed in either group (very common for the PCF 29:13 clone), since the parental Lister 427 strain is known for showing very poor or no SG infections. Altogether, these results suggest that RBP6 is apparently not essential for parasite infection of the tsetse, although it cannot be ruled out that the weak gene knockdown (~50%) may not be enough to lead to a phenotypic change.

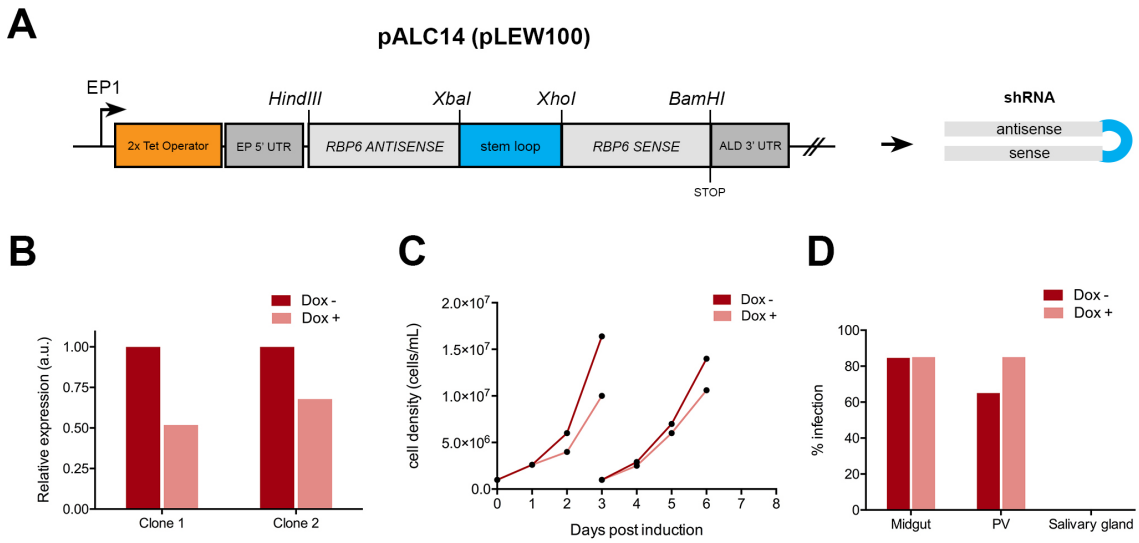


Figure 4.8. RBP6 RNAi knockdown. (A) DNA construct used for the expression of a shRNA to target *rbp6* for RNAi silencing. **(B)** RNA relative expression levels of *rbp6* in two PCF clones. **(C)** Growth curve of mutant PCF (clone 1) either induced RNAi (Dox +) or uninduced (Dox -). **(D)** Infectivity of RBP6 RNAi mutants (Clone 1) in the tsetse midgut, proventriculus and salivary glands (SG).

4.2.8 Identification of other *T. brucei* genes important to progress towards a SG infection and transmission

4.2.8.1 Two *T. b. brucei* strains as a model of study

Having corroborated that RBP6 is involved in life cycle progression, I next tried to identify other genes whose expression is essential for *T. brucei* development in the tsetse. The gene identification was made using a comparative transcriptomics study between two *T. b. brucei* strains with very different tsetse infection phenotypes. The wildtype-like strain TSW-196 (Paindavoine et al., 1986) is pleomorphic and fly-transmissible, while the monomorphic strain Lister 427 29:13, broadly used by the trypanosome community, is unable to progress beyond the tsetse PV and, therefore, it cannot generate SG infections. Thus, these phenotypic differences are likely consequences of changes in gene expression as a result of having been passaged *in vitro* for many years. Comparing whole gene expression between the two strains in PV parasites was the chosen approach to potentially identify expression of genes essential for SG infection.

First, the two *T. brucei* strains under study were fully characterised regarding tsetse infectivity and parasite development through the life cycle. There are many reasons by which the strain Lister 427 29:13 may be unable to progress beyond the tsetse PV. Three major possibilities were considered: 1) parasites are unable to differentiate into EMFs in the PV; 2) PV parasites are unable to migrate to the SG; and 3) parasites are unable to establish a SG infection and/or unable to undergo metacyclogenesis. Tsetse were infected with both strains separately and parasites from MG and PV were collected for analysis at 15 dpi and 30 dpi, and at 30 dpi from SGs (TSW-196 only) (**Figure 4.9-A**). Regarding their capability to infect and develop in the tsetse, both strains can establish ~100% MG infections at 15 dpi, which remain sustained until 30 dpi, as expected. Both strains efficiently infect (~80%) the PV at 15 dpi and this is even increased with time when scoring infection at 30 dpi (~100%). Differentiation to long and short epimastigotes in the PV occurs for both strains (**Figure 4.9-B**). However, and as expected, while TSW-196 establishes normal SG infections at 30 dpi (~22%), infection of SGs by the Lister 427 are strikingly inefficient (<2%) (**Figure 4.9-A**). Importantly, the few SGs that were infected by Lister 427 showed an extremely low number of parasites (<5 MCF cells found per SG pair), compared to the tens of thousands MCF cells found in SGs infected with TSW-196.

Lister 427 MCFs are morphologically indistinguishable to those from TSW-196 (**Figure 4.9-A inset**). However, their motility appears to be compromised, which questions about their viability and whether they would be able to further propagate the infection in a mammalian host.

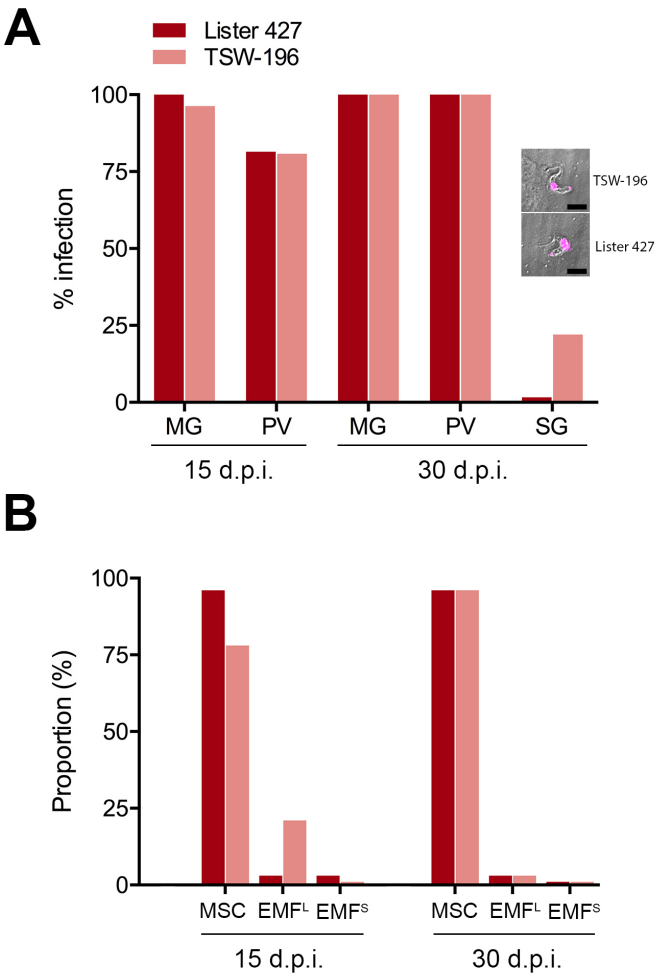


Figure 4.9. Comparison of wildtype *T. b. brucei* strains Lister 427 29:13 and TSW-196. (A) Infectivity rates of tsetse midgut, proventriculus and salivary glands (SG) at 15 and 30 dpi. The inset shows representative MCF cells of both strains stained with DAPI (magenta). Scale bars = 5 μm. **(B)** Proportion (percentage) of mesocyclics (MSC), long epimastigotes (EMF^L) and short epimastigotes (EMF^S) found in PVs either infected with Lister 427 or TSW-196 strains at 15 dpi or 30 dpi.

4.2.8.2 Differential gene expression analysis of *T. brucei* in the tsetse

Tsetse flies were infected with both strains separately and whole gene expression was determined by RNA-seq in PV trypanosomes[3]. A differential expression (DE) analysis revealed 786 genes up-regulated and 867 down-regulated in PV parasites from the 'competent' TSW-196 strain compared with the 'defective' Lister 427 29:13 strain (**Figure 4.10**). Up-regulated TSW-196 genes may reflect defective expression of those in Lister 427, which may account for its inability to progress beyond the PV. Using BLAST2GO, differentially expressed genes were annotated and searched for GO terms. Strikingly, a ~33% of up-regulated genes were related with ribonucleotide binding activity, ~30% with ATP-related processes, and ~13% with biopterin transport (**Figure 4.10**). These putative functions, shared between most of the up-regulated genes, are related to essential cellular processes such as gene expression regulation, energy metabolism and folate import. Folate refers to several essential metabolites needed in the metabolism of nucleic acids, several amino acids and methylation (Fenech, 2012), and both *T. brucei* and tsetse are auxotrophic (Gibson et al., 2016; Sienkiewicz et al., 2008). Thus, both organisms rely on the folate produced by the tsetse bacterial symbionts (Wang et al., 2013). The top 5 up-regulated TSW-196 hits, summarised in **Table 4.1,** are: 1) hypothetical protein (4.9 $\log_2$ fold-change (logFC)), 2) folate transporter (4.4 logFC), 3) glutamate dehydrogenase (4.3 logFC), 4) folate transporter (4.2 logFC), 5) folate transporter (3.7 logFC). Interestingly, *rbp6* was found to be up-regulated in TSW-196, with a logFC of ~2.0 in position 79 (top up-regulated). A defect in *rbp6* expression may compromise cell differentiation of Lister 427 in some way. In addition, introducing *rbp6* in the experimental validation was useful as a positive control for differentiation.

---

[3] Tsetse infections and dissections were done by Dr. Lee Haines and Dr. Alvaro Acosta-Serrano (LSTM, UK). Transcriptomics performed at the Centre for Genomic Research by Dr. Alistair Darby (University of Liverpool, UK). Data analyses done by Ms. Janna Shamsani and Dr. Pegine Walrad (University of York, UK)

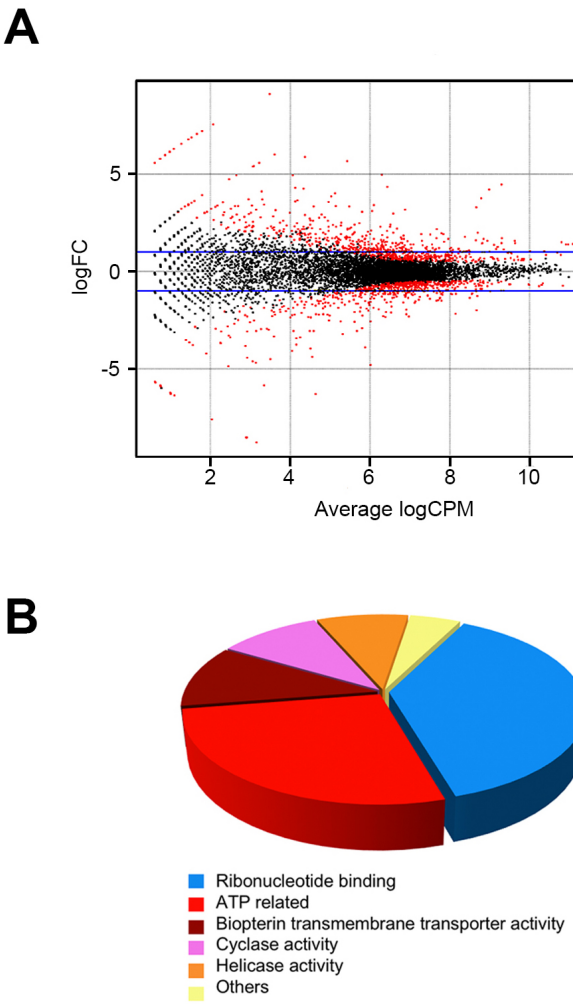


Figure 4.10. Differential expression gene analysis between Lister 427 29:13 and TSW-196 PV trypanosomes at 15 dpi. (A) MA-plot depicting the $\log_2$ expression fold change (logFC) of individual genes compared with the average counts per million (logCPM). Significant hits (p-value ≤0.05) in red and non-significant (p-value >0.05) in black; blue lines set the logFC=2 threshold. **(B)** Pie chart showing the distribution of the major categories of GO functional terms of genes up-regulated in TSW-196 compared to Lister 427.

Table 4.1. Top up-regulated *T. b. brucei* genes in TSW-196 compared to Lister 427 29:13 proventricular trypanosomes. Differential expression analysis performed on the RNA-seq data from *T. brucei* parasites from both strains infecting the tsetse PV at 15 dpi. Gene accession code in TriTrypDB (Gene ID), normalised expression from two biological replicates (1, 2) of TSW-196 (T) and Lister 427 29:13 (L), log_2 fold change T/L (logFC), statistic p-value (p) and protein function.

Gene ID	T1	T2	L1	L2	logFC	p	Function
Tb927.4.4670	558	474	18	12	4.95	3.23E-53	Hypothetical protein
Tb927.8.3630	4428	3810	130	218	4.44	5.19E-79	Folate transporter
Tb927.9.5900	730	386	36	14	4.31	1.52E-24	Glutamate dehydrogenase
Tb927.8.3650	3518	3066	130	200	4.19	2.47E-79	Folate transporter
Tb927.8.3620	2836	2796	116	274	3.74	5.55E-39	Folate transporter
Tb927.3.2390*	1338	1068	338	220	1.94	1.38E-14	RNA-binding 6 (RBP6)

*Tb927.3.2930 occupies the position 79 of up-regulated genes.

4.2.8.3 *In silico* analysis of top candidate proteins

Basic *in silico* predictions were run on the candidate protein sequences to get more information about putative protein features and structure. The Tb927.4.4670 hypothetical protein (483 residues, 54.1 kDa) was predicted to have a zinc finger MYND-like domain (Myeloid, Nervy, and DEAF-1) at C-terminus. This domain, consisting in a cluster of cysteine and histidine residues forming a zinc-binding motif, is not directly related with any particular function (Gross and McGinnis, 1996; Leinhart and Brown, 2011; Spellmon et al., 2015). Its structure was modelled using I-TASSER and found to be a globular protein that is structurally similar to Nup proteins forming the human nuclear pore complex (**Figure 4.11)** (Kosinski et al., 2016). No signal

peptide, transmembrane domain nor GPI-anchor peptide were found and its cellular localisation has not been yet described. The C-terminus was randomly used as the location for epitope tagging for subsequent microscopy analysis (see Section 4.2.8.5).

The paralog genes (Tb927.8.3620, Tb927.8.3630, Tb927.8.3650) encode three isoforms of folate transporters (632 aa, 69.9 kDa), which share a 97.8% of protein sequence identity. In addition, a copy of the gene is found in the BES as the ESAG10, although it has been shown that it is not essential for BSF cells as most of the ESAG proteins (Gadelha et al., 2015). These proteins were recently shown to have folate transport activity (Dewar et al., 2016). Although there is no prediction for cellular localisation, it was shown that ESAG10 localises on the plasma membrane of BSF cells (Gadelha et al., 2015). Membrane localisation is also supported by the presence of 12 predicted transmembrane domains with both free ends facing the cytosolic side. No signal peptides were predicted. The structural model generated in I-TASSER shows a typical transporter barrel shape (**Figure 4.11**) in which the N-terminus seems to be less involved in the overall folding. Thus the N-terminus was found to be the preferred end for protein tagging (see section 4.2.8.5).

The gene Tb927.9.5900 encodes for a glutamate dehydrogenase (992 aa, 112.1 kDa). This enzyme is broadly conserved in eukaryotes as it is responsible for the reversible catalysis of glutamate to form α-ketoglutarate in the mitochondrion (Smith et al., 2017). Hence, it might be a crucial enzyme for cells relying on amino acids as a source of carbon and energy (trypanosomes in the tsetse), because it links the amino acid catabolic pathway with the TCA cycle (Mantilla et al., 2017). In many species, the globular GDH (**Figure 4.11**) adopts a hexamer as quaternary structure by monomer binding through the N-terminus (Son et al., 2015), where a small signal peptide for mitochondrial localisation is predicted. Therefore, the selected tagging site was at protein C-terminus.

Lastly, RBP6 (Tb927.3.2930, 239 aa, 27.1 kDa) was also included in these analyses. No predictions for signal peptides, GPI-anchor, or transmembrane domains were found. Its RNA-binding (AU-rich binding) domain was found to be at the C-terminus, and although its cellular location cannot be inferred from the sequence, it is probably expressed in the cytosol. Protein structure modelling with I-TASSER (**Figures 4.1** and **4.11**) showed that the N-terminal end is more disordered and far

away from the RNA-binding domain at C-terminus, thus suggesting the N-terminal end is the preferred for tagging.

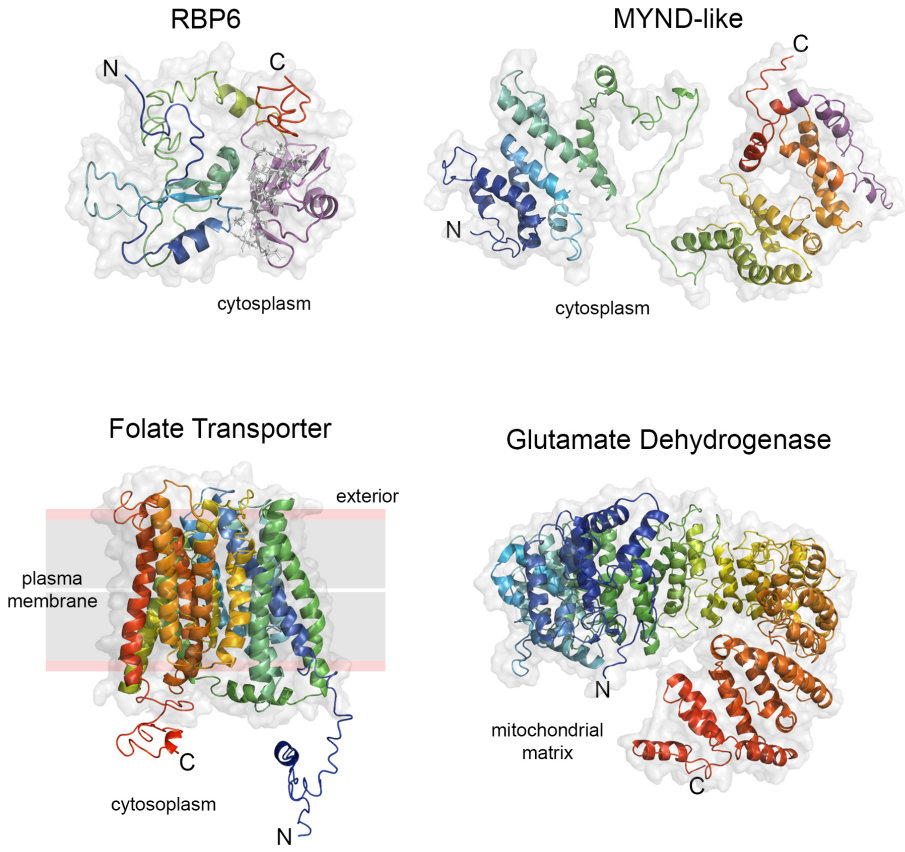


Figure 4.11. 3D models of the candidate proteins. Models generated using I-TASSER from the mature protein sequences of RBP6, the MYND-like hypothetical protein, folate transporter and glutamate dehydrogenase. Molecular surfaces in grey and protein secondary structures in cartoon representation, coloured from blue (N-termini, N) to red (C-termini, C). The RNA-binding domain in RBP6 and the MYND-like domain in the hypothetical protein are coloured in magenta. Proteins represented in their putative sub-cellular location.

4.2.8.4 Assessing expression of the candidate genes

To validate the DE analysis, the RNA relative expression levels of each of the candidate genes were determined using qRT-PCR. This was determined for both TSW-196 and Lister 427 29:13 parasites isolated from the tsetse MG, PV and SG, at

two-time points of infection (15 dpi and 30 dpi) (**Figure 4.12**). Samples from wildtype PCF and cBSF were included. Samples from PCF RBP6 overexpressors either induced or not at 3 dpi were also analysed to investigate whether the RBP6-induced differentiation triggers any candidate gene expression changes.

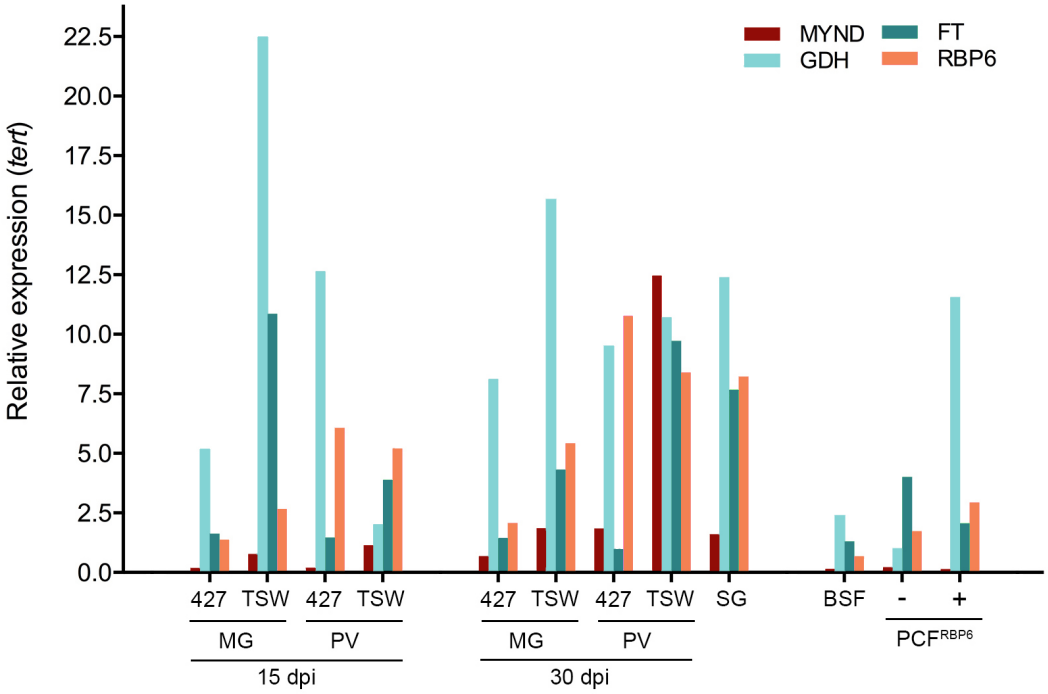


Figure 4.12. Gene expression of the candidate genes throughout the life cycle of *T. b. brucei*. RNA relative expression levels of the hypothetical MYND-like protein (MYND), the folate transporters (FT), glutamate dehydrogenase (GDH) and RNA-binding protein 6 (RBP6), normalised to the expression of *tert*. Parasites from midgut (MG) and proventriculus (PV) at two different time points (15 and 30 dpi) from either TSW-196 or Lister 427 strains; TSW-196 salivary glands (SG). Additional samples from BSF and RBP6 PCF overexpressor cells either induced (+) or uninduced (-) (3dpi).

The MYND-like hypothetical protein was found to have low expression in cBSF, PCF and MG procyclics, and not to be affected by RBP6-induced differentiation (**Figure 4.13**). However, gene expression was found to be consistently higher in TSW-

196 parasites in MG and PV compared with Lister 427 trypanosomes, especially in PV infections (>9-fold up-regulation at 15 dpi and >7-fold at 30 dpi), in accordance with the transcriptomics data. MYND expression achieves its highest in PV infections at 30 dpi, and it is further down-regulated in SG parasites, suggesting it may function in late PV stages.

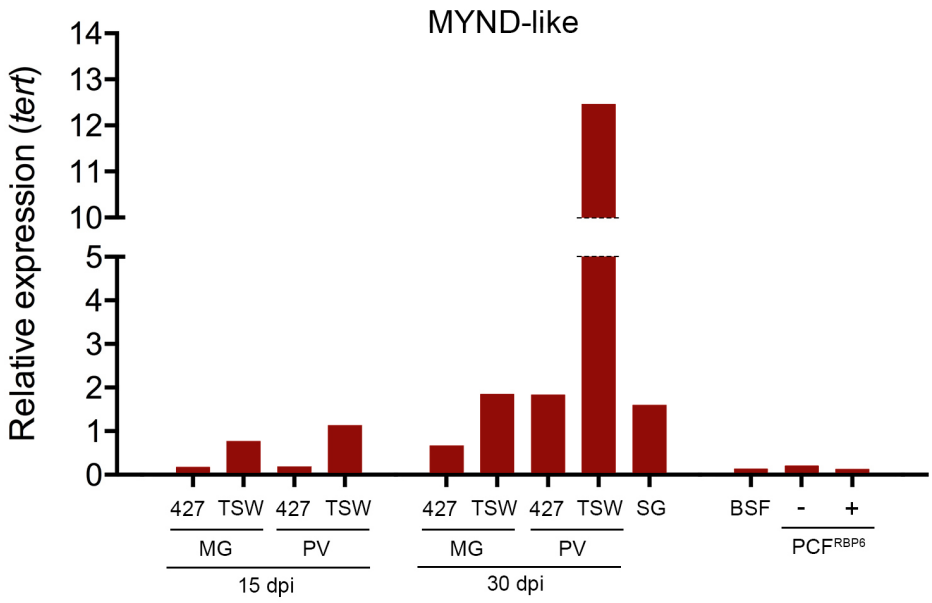


Figure 4.13. Gene expression of the hypothetical MYND-like protein during the life cycle of *T. b. brucei*. RNA relative expression normalised to the expression of *tert*. Same samples as in Figure 4.12.

On the other hand, the folate transporters (expression of three paralogs analysed together) seem to be highly expressed in PV and SG parasites, especially at 30 dpi although high expression was observed in TSW-196 midgut procyclics at 15 dpi (**Figure 4.14**). The expression of the folate transporters is consistently higher in TSW-196 compared with Lister 427, with a ~3 fold and ~11-fold up-regulation in PV trypanosomes at 15 dpi and 30 dpi, respectively. FT expression seems not to be altered upon RBP6 overexpression *in vitro*.

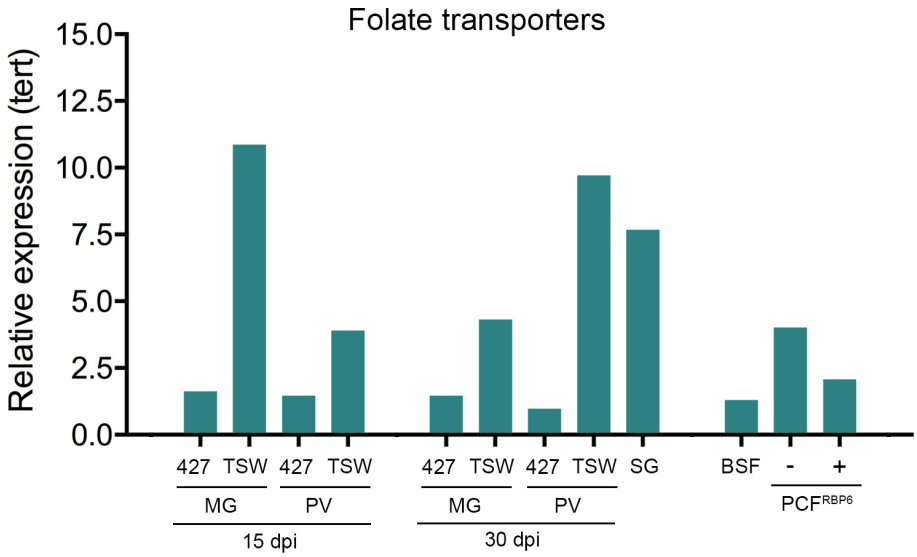


Figure 4.14. Gene expression of the three folate transporters during the life cycle of *T. b. brucei.* RNA relative expression levels normalised to the expression of *tert.* Same samples as in Figure 4.12.

GDH showed the highest overall expression in the tsetse stages compared with the other candidate genes (**Figure 4.15**). No major differences were observed between strains except for TSW-196 MG procyclics (15 dpi) having ~4.5-fold more expression than Lister 427, and this being reverted as a ~6.5-fold down-regulation in PV infections. Interestingly, GDH was the only candidate gene that showed a significant change in expression upon RBP6 induction with a 14-fold up-regulation upon RBP6 induction (3 dpi). This suggests that RBP6 may be directly controlling the expression of GDH or, alternatively, the differentiation cascade of changes triggered by RBP6 did indirectly induce an up-regulation of GDH.

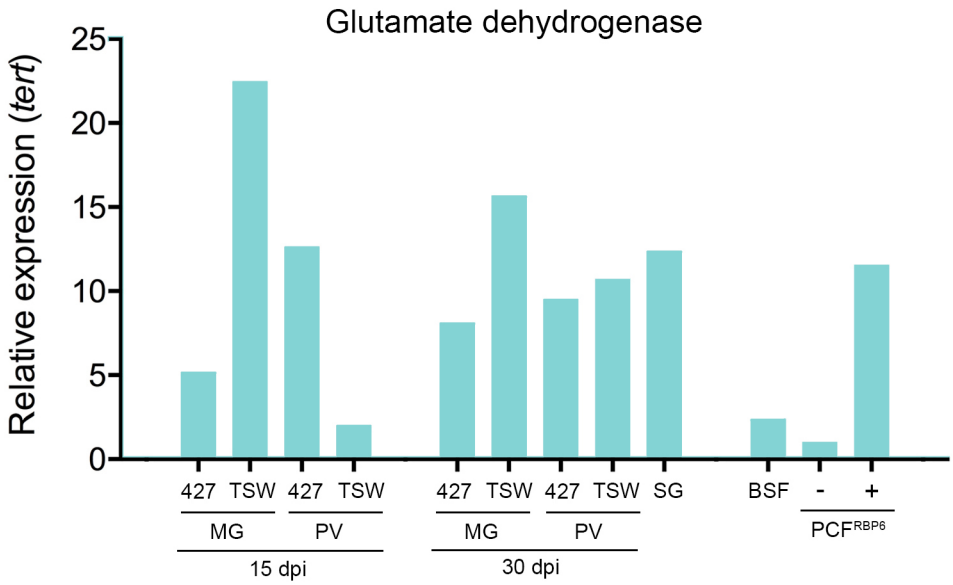


Figure 4.15. Gene expression of the glutamate dehydrogenase throughout the life cycle of *T. b. brucei.* RNA relative expression normalised to the expression of *tert.* Same samples as in Figure 4.12.

The expression of RBP6, although analysed by the same means in section 4.2.5, was here evaluated for the direct comparison of expression levels between *T. brucei* strains (**Figure 4.16**). As expected, RBP6 was found to be up-regulated in PV and SG parasites, but no major differences between strains were detected.

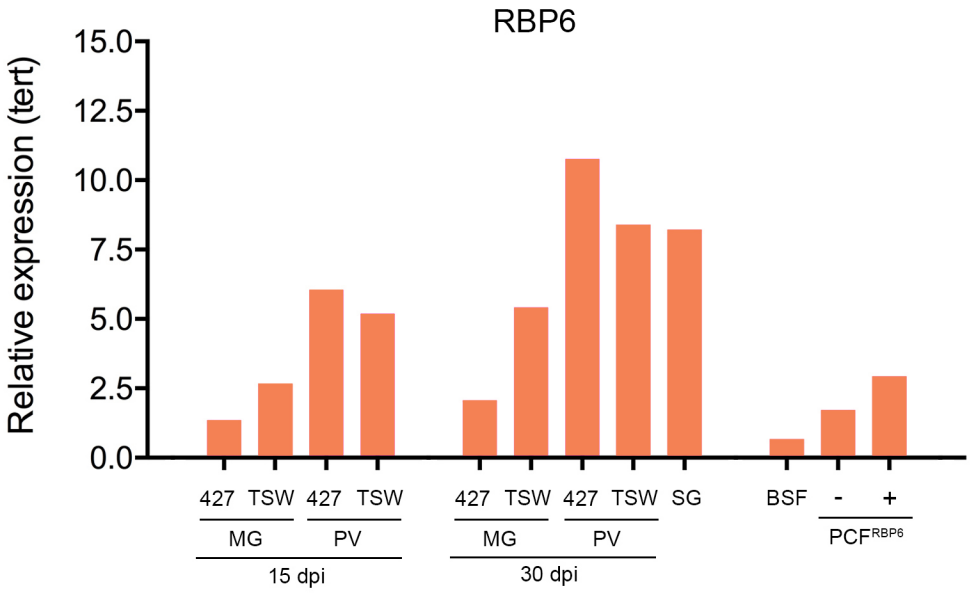


Figure 4.16. Gene expression of the RNA-binding protein 6 during the life cycle of *T. b. brucei.* RNA relative expression levels normalised to the expression of *tert.* Same samples as in Figure 4.12.

4.2.8.5 Analysis of the overexpression of candidate genes in Lister 427

Candidate genes were found to be up-regulated in the competent TSW-196 strain compared to Lister 427 in PV parasites. Thus, overexpressing these genes in Lister 427 (29:13 clone) PV trypanosomes may overcome their deficient expression and potentially lead to rescue the SG infection phenotype. Several DNA constructs were created for the ectopic overexpression of single candidate genes under a strong constitutive rDNA or procyclin promoter and tetracycline-controlled **(Figure 4.17)**. Genes were fused with an HA epitope tag at either N- or C-terminal according to the preferred sites determined in section 4.2.5.8. After transfection, ectopic expression was determined using immunoblotting (**Figure 4.18-A**) and immunofluorescence (**Figure 4.18-B**). Transgenic PCF clones, showing high and homogeneous protein expression, and correct localisation, were further studied.

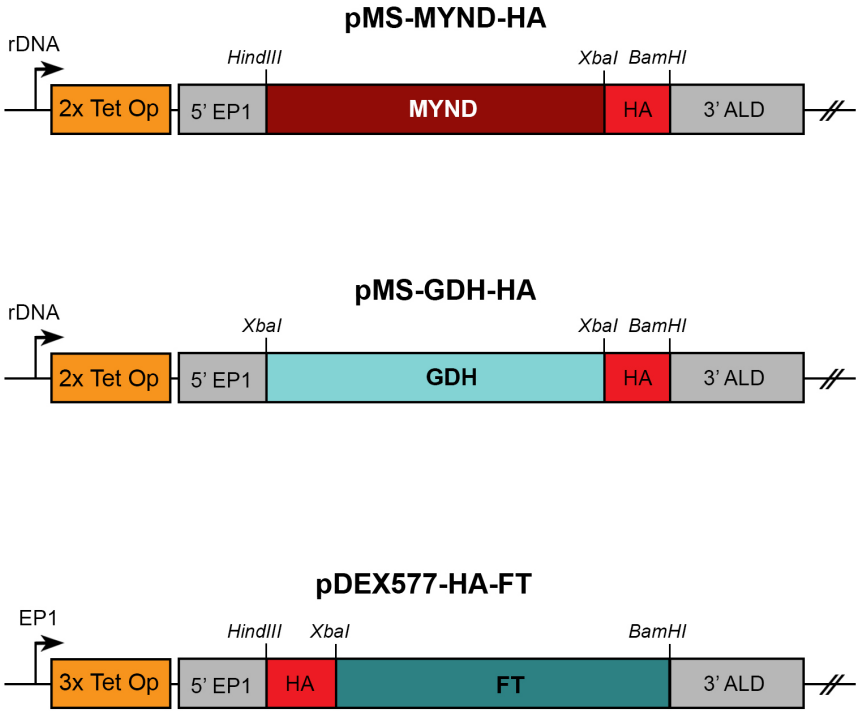


Figure 4.17. DNA constructs for the overexpression of the candidate genes. The plasmids pMS (MYND and GDH) or pDEX577 (FT) were used for the tetracycline-inducible expression of ectopic candidate genes tagged with an HA epitope at either C- (MYND, GDH) or N- terminus (FT). All transcripts under the regulation of the EP1 procyclin 5' UTR and the aldolase 3' UTR.

The hypothetical protein MYND-like was found to be expressed in induced cells by immunoblotting with an apparent molecular size of ~50 kDa (54 kDa expected) and was observed to localise in the cytoplasm (although excluded from the nucleus) by immunofluorescence (**Figure 4.18**). MYND has not been endo-tagged in TrypTag for comparison. In addition, the folate transporter (Tb927.8.3630) localised evenly on the surface of PCF, consistent with its location as a surface transporter (Gadelha et al., 2015). Interestingly, TrypTag localises this protein in the cytoplasm, which is unlikely for a transporter with 12 transmembrane domains. However, it could not be detected by immunoblotting, possibly due to a solubilisation problem common in proteins containing several hydrophobic transmembrane domains. Lastly, the GDH fusion protein was detected by immunoblotting as migrating with an apparent molecular mass of ~100 kDa (112 kDa expected). Moreover, the protein localises in

the mitochondrion as shown by co-localisation with the mitochondrial marker, MitoTracker (Mantilla et al., 2017). In TrypTag, GDH only localises in the mitochondrion when tagged at C-terminus, although it also seems to accumulate in the kDNA, probably as a consequence of being fused to mNeonGreen.

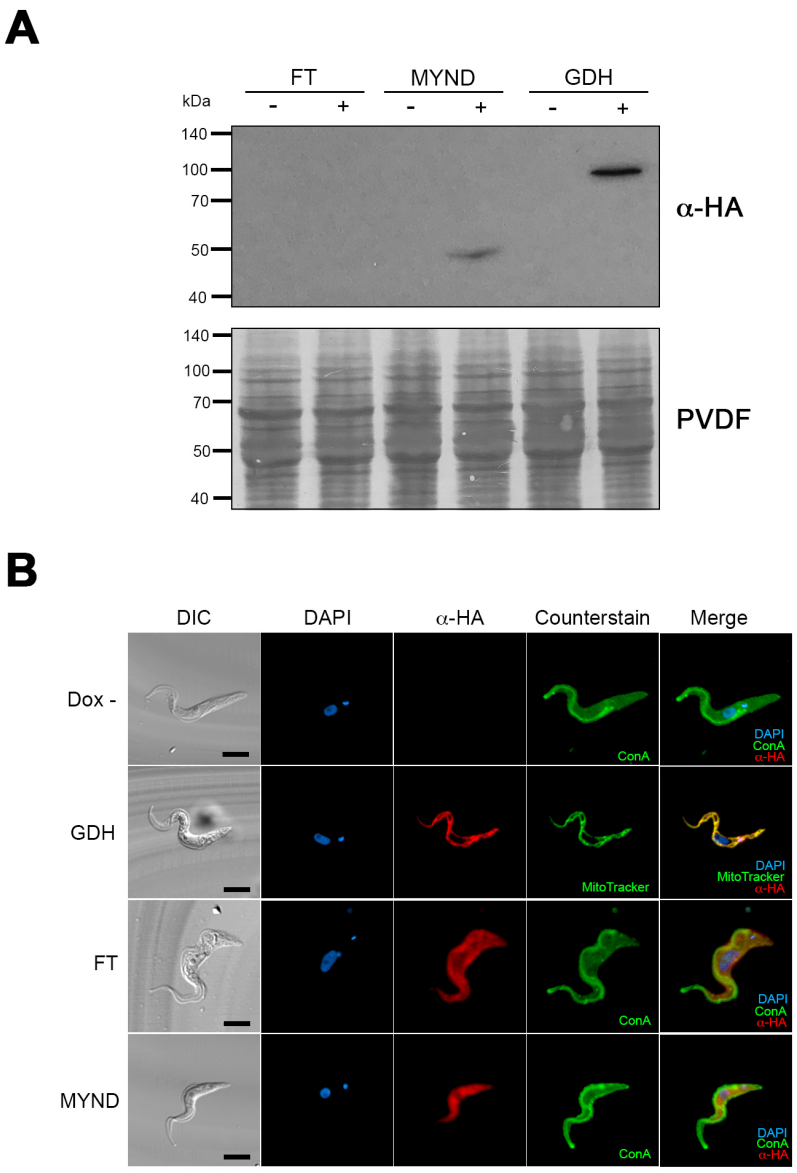


Figure 4.18. Cellular localisation of ectopic candidate proteins. Ectopic GDH, FT and MYND localised by immunostaining in induced (Dox +) cells using DAPI (blue), probed with anti-HA (red), and co-stained (green) with either concanavalin A (cell surface) or MitoTracker (mitochondrion). DIC and fluorescence merged images shown. Scale bars = 5 μm.

4.2.8.6 Infectivity of PCF overexpressing candidate genes in the tsetse

After a basic characterisation of the PCF Lister 427 cell lines overexpressing several of the candidate genes, these cells were used to infect tsetse in order to determine whether any of these genes is able to rescue the SG deficient infectivity. Teneral tsetse were infected with equal amounts of parasites that were either previously incubated with Dox for 2 days (pre-induced) or with uninduced cells (as negative controls). Flies were dissected at 30 dpi and the infection levels scored in the MG, PV and SG (**Figure 4.19**). Remarkably, only the MYND-like protein succeeded to rescue SG infectivity in Lister 427 at rates equivalent to that in wild type strains (~13%). No major differences were observed in MG and PV infection rates.

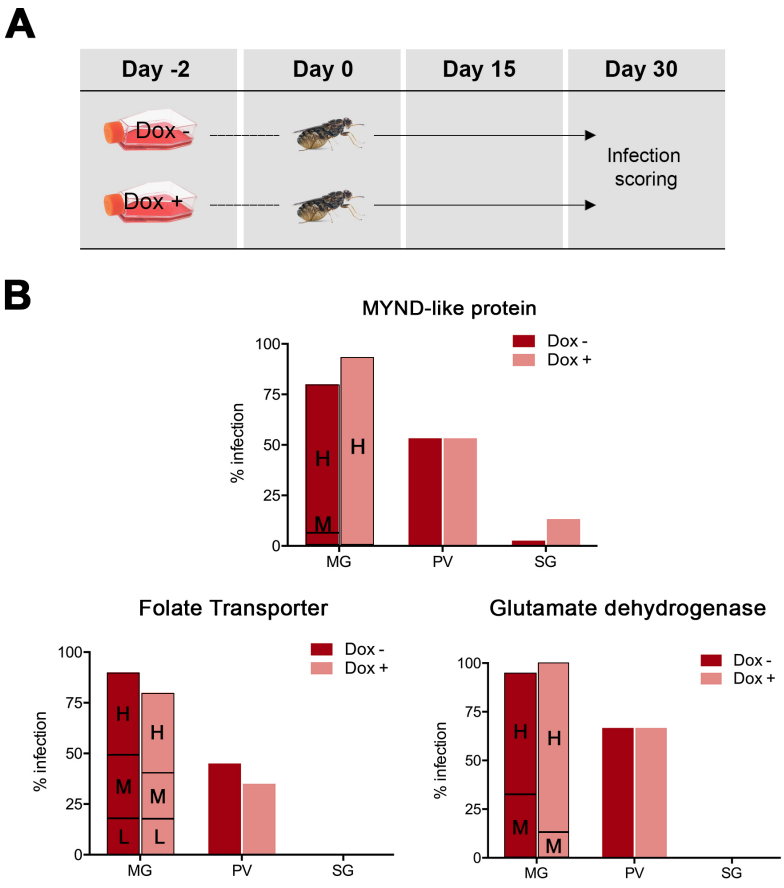


Figure 4.19. *In vivo* phenotype of PCF overexpressing individual candidate genes. (A) Experimental design of infections. **(B)** Infectivity (percentage) of the overexpressors in the tsetse MG, PV and SG. MG infection levels are noted as high (H), medium (M), or low (L).

4.3 Discussion

Studies on *T. brucei* SG stages are scarce compared to those in BSF or PCF. The main reason is the ability to culture these parasite stages *in vitro* so an unlimited number of homogeneous cells can be obtained. In contrast, SG trypanosomes cannot be cultured *in vitro* and so they need to be extracted from infected tsetse. This process is time-consuming and yields a very reduced number of cells mixed with SG debris and saliva, thus making most downstream experiments extremely challenging or unfeasible. For instance, this applies to the studies on MISP and BARP presented in Chapter 3. Hence, a tsetse-independent alternative method to obtain *T. brucei* SG forms would be ideal to overcome these difficulties. In this Chapter, I present the setup of a published method to obtain MCF *in vitro* (Kolev et al., 2012) with a more detailed characterisation and with special focus on the usage of these cells in MISP and BARP studies. In addition, in this chapter I characterised a few genes identified as potentially essential for the parasite to become transmissible in the tsetse SG.

4.3.1 Overexpression of trypanosomal RBP6: a tool to study *T. brucei* development *in vitro*

Since the original report on the role and properties of *T. brucei* RBP6 protein (Kolev et al., 2012), surprisingly, no work has been published using the same system by other groups. After having generated three independent RBP6 overexpressor lines with different characteristics (e.g. expression levels and tags), it seems unlikely that a lack of reproducibility is responsible for this. The new three lines of RBP6 overexpressors generated in PCF responded well to doxycycline induction of the transgene and differentiated to EMF and MCF although with different dynamics. The cell line used to carry out the full characterisation had the RBP6 fused to three HA epitopes and a sfGFP reporter, which greatly eased the selection process without compromising functionality. Upon induction, despite the population being clonal and expressing the ectopic sfGFP-tagged RBP6, a few cells start the differentiation process to EMF. After a few days, MCF were generated from EMFs as it naturally happens in the tsetse. While MCF peaked at 3 dpi, EMF are found to be present during the entire time course of the experiment (10 days). However, it appears that most PCF do not differentiate upon addition of Dox, and that replicating EMF eventually generate MCF, although it cannot be ruled out that PCF population may

differentiate at variable times. Unfortunately, this could only be figured out if individual cells could be isolated and tracked over time, for instance, using a microfluidics device with multiple chambers. One novel finding in all these new RBP6 cells is that upon induction, dividing BSF-like trypomastigotes also appear in culture after 3 dpi. Thus, RBP6 may also promote the differentiation from MCF to BSF. Although these BSF-like parasites could not be maintained in culture, perhaps because they rapidly differentiated to PCF or the culture conditions were not favourable, they have been shown to be replicative and significantly bigger in size than MCF. In addition, these forms also express CRD epitopes but not EP-procyclins, all characteristic of BSFs. MCF in contrast, do not replicate (Matthews, 2005). Further immunodetection studies for the identification of the transferrin receptor exclusively expressed by BSFs (Tiengwe et al., 2017), for instance, could corroborate the identity of these BSFs.

Besides generating different parasite stages, induced RBP6 overexpressors also experience anomalies in growth. It can be observed that right after induction, cells grow slower than the uninduced controls at all times. Changes in growth rate, however, are not stable but greatly fluctuate, perhaps due to two main reasons. First, cell density must have an impact in the growth rates as cells divide faster during exponential phase (10^6 to 10^7 approximately) than in stationary phase (>10^7). Second, cell differentiation is likely to slow cell growth rates because of the number of cells undergoing differentiation.

The setup of the RBP6 overexpression system aimed to produce EMF and MCF for the study of MISP and BARP at a new level. It has been shown that these cells express MISP as in wildtype SG stages (Chapter 3), although the presence of BARP (Urwyler et al., 2007) could not be confirmed. In addition, the ion-exchange purification method (Fox et al., 1986) allows the enrichment of MCF and BSF-like forms from large volumes of culture. Although in this Chapter no studies on MISP and BARP were performed using the *in vitro*-differentiated cells, the production, purification and characterisation methods have been established for future work.

Although it has been shown that RBP6 promotes differentiation *in vitro*, the biological role of RBP6 remains unclear. Expression analyses confirmed that *rbp6* is up-regulated in PV and SG trypanosomes, whereas it is negligible in PF and BSF (Christiano et al., 2017; Savage et al., 2016). This suggests it plays an important role in the development of late stages in the tsetse. In an attempt to validate this, the

knockdown of RBP6 did not compromise parasite development in the tsetse as expected. However, knockdown efficiency in several cell lines was proved to be low (perhaps because of the potential essentiality of RBP6) and remaining RBP6 may be able to support normal functionality. A conditional-null RBP6 mutant will be the ultimate tool to unequivocally determine essentiality of RBP6 and its biological role in the tsetse. Alternatively, RBP6 was overexpressed to understand the role of differentiation promotion in the tsetse, in a strain with potential differentiation defects. As seen *in vitro,* differentiation dynamics upon RBP6 induction are fast and complex. To translate and adapt these phenomena into the tsetse, two induction timing approaches were followed to either resemble the natural differentiation process in the PV at 15 dpi, or to provide the tsetse with a mixture of parasite stages, expecting an early establishment of PV or SG infection. Both approaches failed to rescue SG infectivity of this mutant strain, although it cannot be ruled out that induction timing may have not been accurate enough.

In summary, this chapter section has optimised and better defined an *in vitro* tool to generate *T. brucei* stages that resemble those in the tsetse SG for the study of MISP, among other proteins. It has also provided new insights into the biological role of RBP6 *in vivo*.

4.3.2 Gene expression potentially important for *T. brucei* transmission

This chapter has also aimed to identify genes whose expression is important for *T. brucei* to establish an infection in the tsetse SGs. This is essential for metacyclogenesis and thus completing the transmission cycle. The study was based on the comparison of two *T. brucei* strains which, in principle, are genomically equivalent, but present a distinctive infection phenotype in the tsetse with regards migration or colonisation of SGs. A detailed characterisation showed that the impeded (Lister 427 29:13) strain undergoes a normal development in the PV compared to TSW-196 and it is even capable of generating SG infections, although inefficiently. This suggests that these parasites have a defect in either differentiation after the PV stages, migration to the SG or establishment of an infection in the SG, any of each potentially explained by the differential expression of key parasite genes, as suggested by transcriptomics analysis. The top up-regulated genes in proventricular trypanosomes from the competent strain were a hypothetical protein with a MYND-

like domain, three folate transporters and the glutamate dehydrogenase. These genes were all found to be developmentally regulated and up-regulated in late PV and SG infections, suggesting their products are likely to carry out their function in those parasite stages. It was hypothesised that genes whose expression was down-regulated by the impaired strain in the PV, might be responsible for its low infectivity in SGs. However, with the exception of the MYND-like protein, overexpression of any of the other candidate genes did not rescue the SG infection phenotype. There are multiple hypotheses to consider: 1) none of these genes (except MYND) is directly responsible for the infectivity defect (unlikely because of the ability of RBP6 to promote parasite development), 2) the defective phenotype is due to a combination of multiple genes, 3) the phenotype could be due to a gene or combination of genes up-regulated in the impeded strain, or 4) the defective phenotype is not a direct cause of changes in gene expression but due to physical factors such as motility within the vector, which may be more challenging to study.

As mentioned above, the overexpression of the MYND-like candidate protein was able to rescue the SG infection phenotype yielding infection rates equivalent to that in wildtype strains. In addition, the overexpression of the protein in PCF *in vitro* triggered the formation of EMF (~10%) stages morphologically equivalent to those generated in the RBP6 *in vitro* system (preliminary observation). Although proteins with the MYND domain have been described to have RNA-binding properties (Spellmon et al., 2015), it cannot yet be assumed that this protein is a putative RBP, since other MYND-related functional properties such as protein-binding may also lead to cell differentiation by undetermined mechanism. Future studies using the cell line here created including RNA-seq, immunoprecipitation and gene knockout, could shed more light into the putative role of the MYND-like protein.

In summary, this chapter section has described a comparative study between two *T. brucei* strains to identify genes whose expression may be important for the life cycle completion of *T. b. brucei* in the tsetse vector. Four candidate genes were tested by ectopic protein expression in mutant cell lines, and assessment of their *in vitro* and *in vivo* phenotypes. Only the MYND-like candidate protein, whose function remains undetermined, promoted SG infection and therefore should be further investigated.

CHAPTER 5

Studies on the *N*-glycosylation of *Trypanosoma brucei* tsetse stages

5.1 Introduction

The glycocalyx of *T. brucei* has been extensively studied mainly in two life stages, i.e. BSF and PCF. The cell surface of these two stages is known to be densely coated by two developmentally regulated major surface glycoproteins: VSGs and procyclins, respectively, which can be both *N*-glycosylated (Ferguson, 1997). Overall, the function of these glycosylations remains unclear. It has been shown that VSG *N*-glycans play a role in the dense packing of the VSG coat (Mehlert et al., 2002), whereas glycosylation of EP-procyclins may be important (but not essential) for parasite colonisation of the tsetse midgut (Vassella et al., 2009). So far, nothing is known about the structures and function(s) of surface *N*-glycans from *T. brucei* EMF and MCF. It was hypothesised that, since *N*-glycans are post-translationally added to proteins, the ones found as part of the MCF glycocalyx could be invariable antigens, independent on the type of VSG polypeptide expressed, and thus potentially exploitable for transmission-blocking approaches.

In *T. brucei*, *N*-glycan structures vary from complex-type *N*-glycans mostly (containing poly-acetyllactosamine or αGal terminal residues) in VSGs to high mannose *N*-glycans on EP-procyclins ($Man_5GlcNAc_2$) (Ferguson, 1997; Izquierdo et al., 2009a). *N*-glycan structures are not directly encoded in the glycoprotein genes, but they are the result of multiple and complex interactions between glycosylation enzymes, donor substrates and acceptor proteins, among others (Varki et al., 2015). A key enzyme in the *N*-glycosylation pathway is the oligosaccharyltransferase (OST), which in *T. brucei* is composed by a single subunit (STT3) and whose function is to attach *N*-glycan precursors to nascent glycoproteins within the ER lumen (Izquierdo et al., 2009b). *T. brucei* has three paralog genes encoding for three STT3 isoforms (A-C), which appear to be developmentally regulated. They have different affinities for glycan substrates and acceptor sequences which have a direct impact in the resultant *N*-glycan structure (Izquierdo et al., 2009b). The expression of STT3 isoforms has been determined in BSF and PCF, but it remains unknown which ones are preferentially expressed in the tsetse stages.

In this chapter, I describe the expression profile of the three *stt3* paralogs throughout the entire life cycle of *T. brucei*, including EMF and MCF, which provides new insights into the potential types of *N*-glycan structures found on the surface of these parasite stages. I also attempted to obtain structural information on the *N*-glycans expressed by MCFs. Lastly, functional studies were performed to pinpoint the essentiality of *N*-glycosylation in all developmental stages of *T. brucei* in the tsetse.

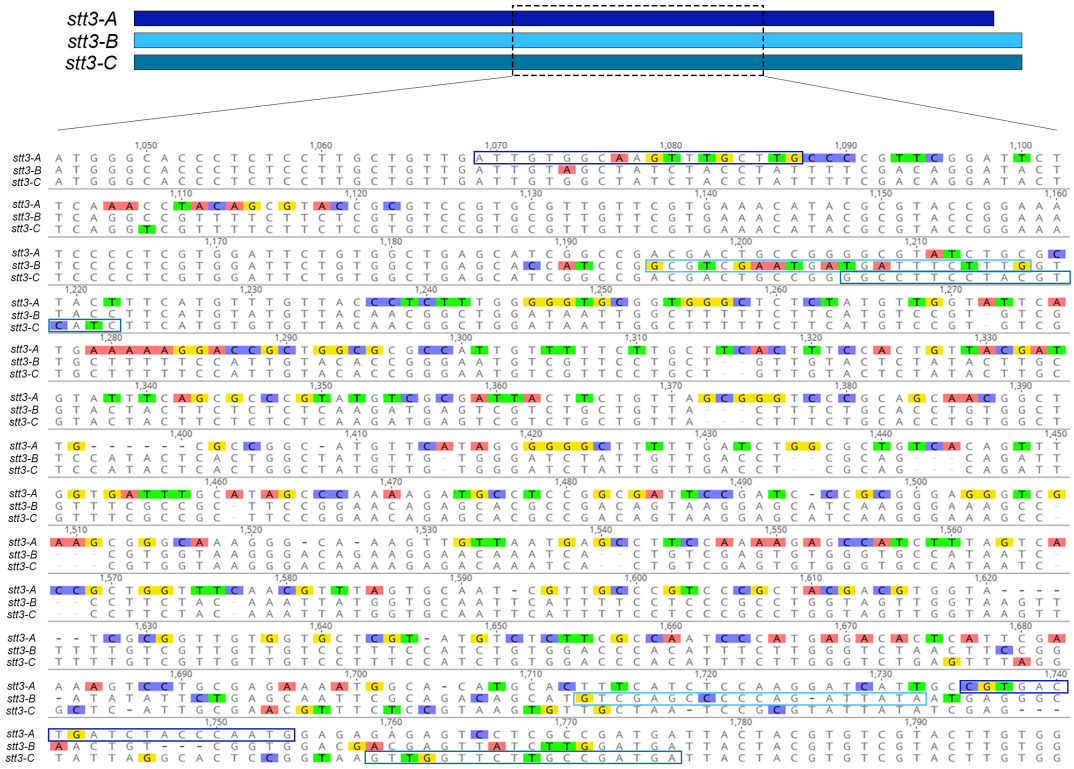
5.2 Results

5.2.1 Developmental expression of *stt3* paralogs

Understanding how *T. brucei* regulates the expression of the three *stt3* paralog genes during its life cycle is key to determine the structures and functions of parasite *N*-glycans. However, these studies have only been conducted in cultured parasites (i.e. BSF and PCF), so the expression pattern of *stt3* in the tsetse stages remains to be investigated. The differential RNA expression levels of all the *stt3* paralogs (*stt3-A, stt3-B, stt3-C*) were determined from cDNA in all *T. brucei* life stages: cBSF, PCF, MG procyclics, PV forms and SG forms. Since the DNA sequence identity shared between paralogs is very high (~84% between *stt3-A*/B, ~96% between *stt3-B/C*, and ~84% between *stt3-A/C*), it was virtually unfeasible to use quantitative real-time RT-PCR. Semi-quantitative RT-PCR was the method of choice due to its increased flexibility in design. Primers were designed for the specific and unequivocal amplification of individual paralogs exploiting the gene regions with highest sequence diversity (**Figure 5.1-A**), producing amplicons of different sizes (686 bp for *stt3-A*, 538 bp for *stt3-B* and 571 bp for *stt3-C*). PCR reactions were stopped at exponential phase and the products were analysed using gel electrophoresis.

The analyses first confirmed what has been reported in BSF and PCF (Izquierdo et al., 2009b). BSFs express both *stt3-A* and *stt3-B* at similar rates, while PCFs primarily express *stt3-B* and little *stt3-A.* In addition, little expression of *stt3-C* was detected in PCF and in BSF it was found to be equivalent to that of *stt3-A*. When analysing the expression in the tsetse stages, little *stt3-C* was detected in MG parasites along with *stt3-B*, which was the only expressed paralog in PV and SG forms.

A



B

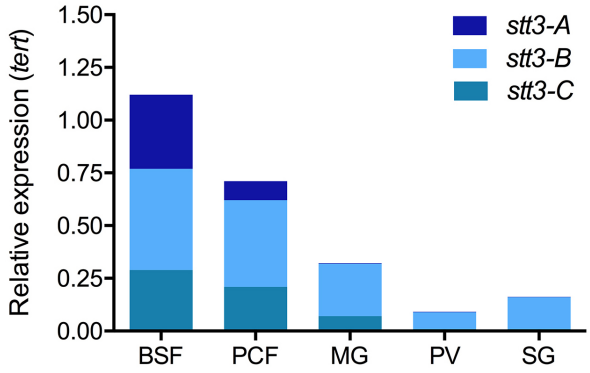


Figure 5.1. Developmental expression of the oligosaccharyltransferase paralogs in *T. brucei*. (A) Schematic of the three *stt3* paralogs highlighting the domain with least sequence identity (top). Multiple sequence alignment of the DNA coding sequences of *stt3-A*, *stt3-B*, and *stt3-C*. Forward and reverse primers designed for the detection of the individual paralogs by RT-PCR are shown in boxes. **(B)** RNA relative expression levels of *stt3* paralogs throughout the life cycle of *T. brucei* (strain AnTat 1.1 90:13), normalised to the expression of *tert.*

5.2.2 Functional studies on *N*-glycosylation

5.2.2.1 Attempts to generating *stt3* and *rft1* conditional-null mutants

To study the role(s) and essentiality of *T. brucei N*-glycosylation during development in the tsetse, several *T. brucei* mutant cell lines defective in *N*-glycosylation were created by targeting two key enzymes in the *N*-glycosylation pathway: the oligosaccharyltransferase STT3 and the putative flippase RFT1 (Helenius et al., 2002). To generate a conditional-null mutant, I tried to knock out the two alleles of these genes by replacing the coding sequence with an antibiotic resistance marker, in separate rounds of transfection. Both 5' and 3' UTR sequences of the genes were used to target the insertion of the markers (**Figure 5.2**). Lastly, an ectopic HA-tagged copy (addback) of the knocked-out genes was introduced in an untranslated region of the genome (intergenic rRNA), whose expression is tetracycline-inducible. This versatile system allows to rescue the wildtype phenotype by inducing the addback with tetracycline in a KO-like cell line (-Tet), or to observe the KO phenotype in a wildtype-like cell line (+Tet) by removing tetracycline from the culture. There are three *stt3* paralog genes (*Tb927.5.890, Tb927.5.900, Tb927.5.910*) and one pseudogene (*Tb927.5.880*) in chromosome 5 in an array disposition. The four genes were knocked-out by replacing the entire array with the antibiotic marker (**Figure 5.2**). In contrast, there is a single *rft1* gene which was normally targeted for replacement (**Figure 5.2**). While knocking out *stt3* alleles did not present much difficulties, the single KO *rft1* cells were refractory to get the second allele removed in several attempts. To fix the issue, single KO *rft1* cells were provided with the *rft1* addback whose expression was kept permanently on and indeed allowed the knockout of the second allele.

A PCR approach was designed to genotype the generated mutant cell lines for the double KO of all *stt3* and *rft1*, with an addback gene. Multiple primers were designed to target unique regions to unequivocally detect individual insertions and insertion sites (**Figure 5.3**). Results showed that all cell lines had successfully inserted the puromycin and blasticidin resistance markers in the expected loci and, in RFT1 lines, the ectopic addback was inserted in the genome too. However, all cell lines also showed that, despite having removed the two native alleles, at least one copy of the genes had remained in the genome, likely as a result of a gene duplication

event during the process (Emes and Yang, 2008). Consequently, none of the created mutants was found to be useful in the end for further studies on *N*-glycosylation.

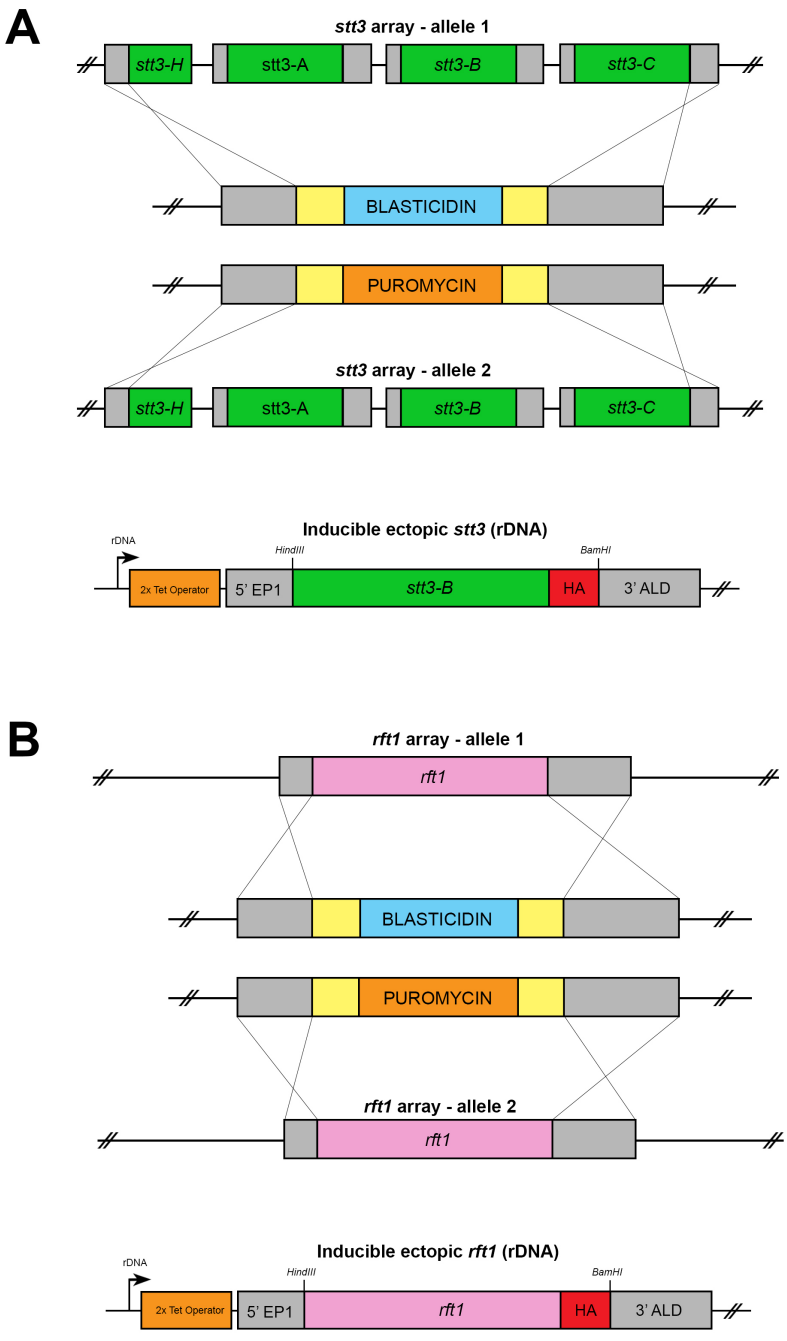


Figure 5.2. DNA constructs for the creation of conditional-null *stt3* and *rft1* mutant trypanosomes. (A) Genetic approach for the knockout of the STT3 array and the addition of an ectopic HA-tagged *stt3-B* in *T. b. b.* AnTat 1.1 90:13. **(B)** Approach for the double knockout of RFT1 and the introduction of an ectopic RFT1 HA-tagged.

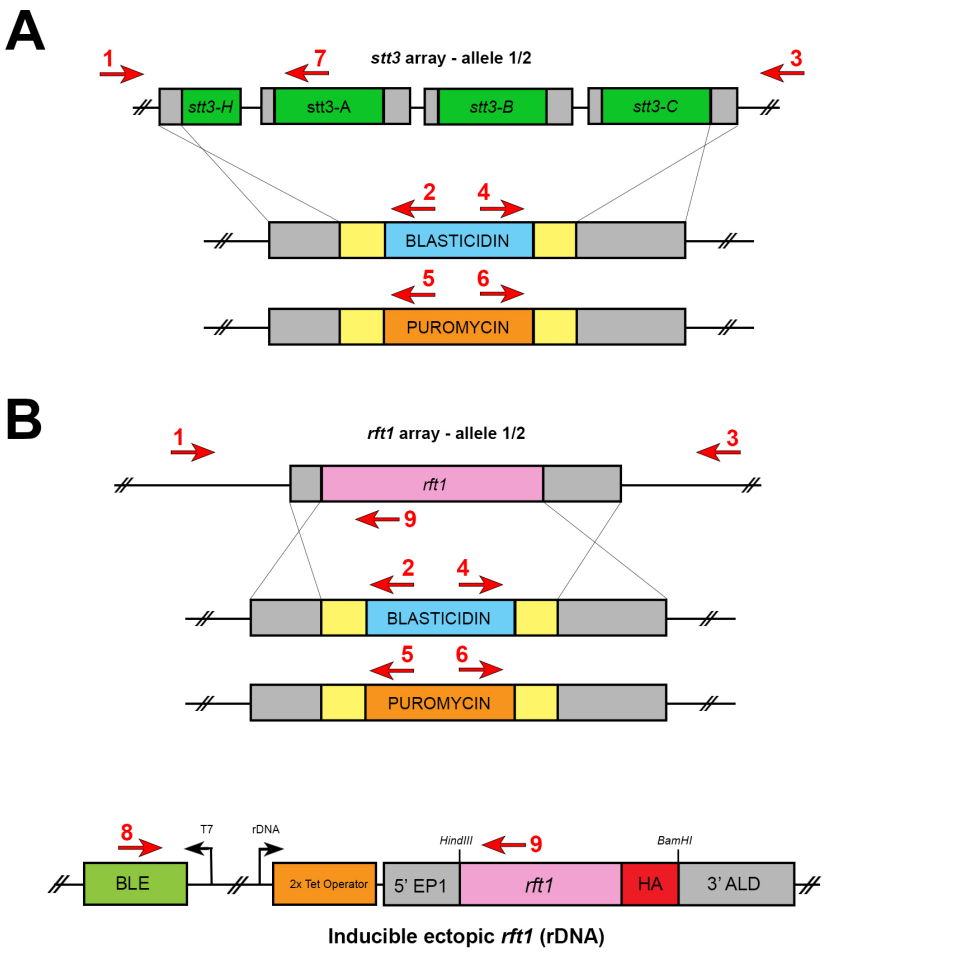


Figure 5.3. Genotypic analysis of conditional-null *stt3* and *rft1*mutants. Primer design (red arrows) to identify the insertions for the double knockout of STT3 **(A)** and RFT1 with addback addition **(B). (C)** Electrophoresis results of mutant genotyping; parental (W), single KO (1) and double KO (2).

5.2.2.2 Creation of *stt3* RNAi knockdown mutants

Having failed in generating the *stt3* and *rft1* conditional-null mutants, I then used RNAi to knockdown the expression of the *stt3s*. RNAi, although simpler and faster, it is not as definitive as the conditional-null knockout approach. A DNA construct was engineered for the overexpression of dsRNA to target the transcripts from the three *stt3* paralogs (**Figure 5.4-A**). A 537 bp unique sequence fully conserved in *stt3H, stt3A, stt3B* and *stt3C*, was amplified by PCR and cloned in both sense and antisense directions flanking a stem loop sequence, under the expression of an EP-1 procyclin promoter and tetracycline-inducible. The construct was transfected in *T. brucei* PCF resulting in clonal cell lines expressing the short-hairpin dsRNA upon doxycycline induction to silence *stt3* expression. The knockdown levels were first determined by RT-PCR using universal primers to detect all *stt3* transcripts (**Figure 5.4-B**). Several clones were screened and the one with the strongest downregulation of *stt3* (~65%) was selected for further experiments.

5.2.2.3 Characterisation of s*tt3* RNAi knockdown mutants

First, a phenotypic experiment was performed to measure the *N*-glycan levels on the cell surface and the effect of the *stt3* RNAi on it. Uninduced and induced PCF were incubated with a fluorescent ConA which binds to high-mannose *N*-glycans attached to surface EP-procyclins (Acosta-Serrano et al., 2000). Thus, the levels of surface *N*-glycosylation were directly correlated with fluorescence intensity. To test the specificity of the ConA, the lectin was pre-incubated with sugar inhibitors (αMe-Man and αMe-Glc) and to test the sensitivity, PCF were pre-incubated with tunicamycin, a known inhibitor of eukaryotic *N*-glycosylation (Takatsuki et al., 1971). Cell fluorescence was analysed using Flow-Activated Cell Sorting (FACS) (**Figure 5.4-C**). The lectin was found to be specific since cells stained with pre-inhibited ConA displayed fluorescence levels equivalent to unstained negative control cells. In addition, the pre-incubation with tunicamycin resulted in a partial inhibition of *N*-glycosylation as cells showed increased fluorescence compared with unstained cells, but a 10-fold reduction compared with non-treated stained cells. When analysing the binding of ConA to *stt3* RNAi mutants, uninduced cells showed fluorescence levels equivalent to that in wildtype stained cells. However, when the knockdown was

induced, cells displayed ~5-fold less fluorescence than uninduced cells, which proves the *stt3* knockdown led to a partial reduction in surface *N*-glycosylation.

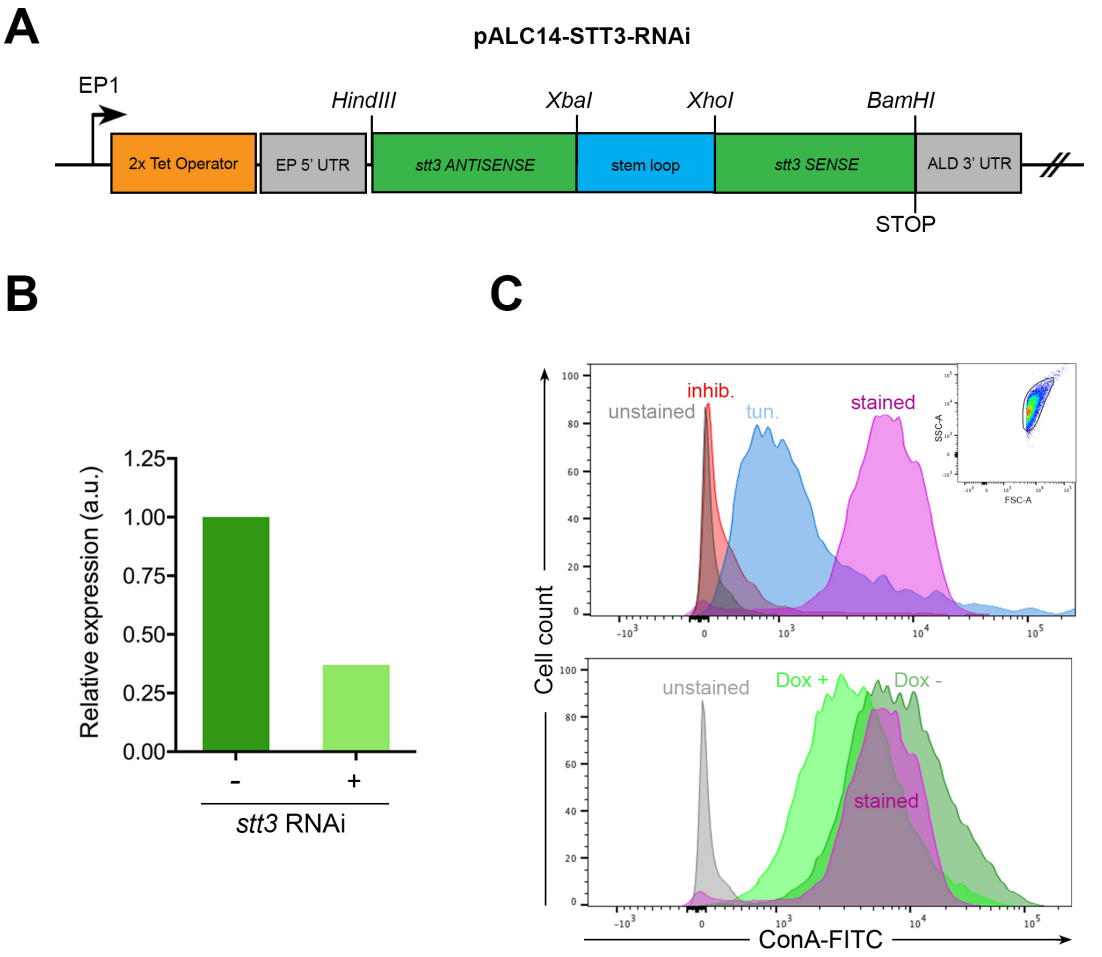


Figure 5.4. Characterisation of *stt3* RNAi AnTat 1.1 90:13. (A) DNA construct designed for the expression of a shRNA to downregulate all *stt3* transcripts through RNAi. **(B)** RNA relative expression levels of all *stt3* in RNAi cells either uninduced or induced state. **(C)** FACS analyses on *stt3* RNAi cells. Top panel shows a histogram for ConA-FITC intensity of control wildtype cells tunicamycin-treated (tun.) and ConA inhibited (inhin.) compared to stained and unstained cells. Top panel inset shows a representative dot plot (FSC-A vs SSC-A) of the population of cells analysed (gated). Bottom panel shows a histogram for ConA-FITC intensity of RNAi cells either induced (Dox +) or uninduced (Dox -) compared to wildtype stained and unstained cells.

A different approach was used in addition to characterise the phenotype of *stt3* RNAi cells. Both immuno- and lectin blottings were performed to analyse the reduction of *N*-glycosylation in RNAi cells at a more detailed level. Procyclins from both uninduced and induced cells (and tunicamycin treated ones) were extracted with organic solvents and their migration analysed by western blotting using specific anti-EP antibodies (Butikofer et al., 1997). The blotting revealed that procyclins from all samples migrated as single broad bands with an apparent molecular mass of ~50 kDa (**Figure 5.5-A**). However, procyclins from tunicamycin-treated cells showed a slightly faster migration, likely as a result of the loss of the *N*-glycans. Furthermore, procyclins from RNAi induced cells revealed the broadest band compared to both untreated wildtype and RNAi uninduced cells, probably as a result of having a mixture of glycosylated and non-glycosylated procyclins as RNAi only produced a partial knockdown. The same procyclin extracts were then analysed by ConA-blotting (**Figure 5.5-B**). Although less specific than anti-EP, the lectin reveals not only the glycosylated procyclins, but also the presence of other *N*-glycosylated proteins (likely contaminants in the procyclin preparations). Untreated wildtype cells showed a broad range of *N*-glycoproteins while tunicamycin-treated cells showed a massive reduction in signal, as expected due to overall *N*-glycosylation inhibition. The *stt3* RNAi induced cells showed an equivalent reduction in detection, which is comparable to that from tunicamycin-treated cells. This shows that tunicamycin and *stt3* RNAi not only affected the *N*-glycosylation of EP-procyclins, but had a general cellular effect. The results were also compared with the binding of ConA to chicken albumin (as a positive control for detection), as well as wildtype procyclins either treated or not with the glycosidase PNGase F, an enzyme that cleaves off *N*-glycans from proteins. Strikingly, there seems to be at least one protein resistant to PNGase F cleavage, but its identity was not further confirmed by mass spectrometry.

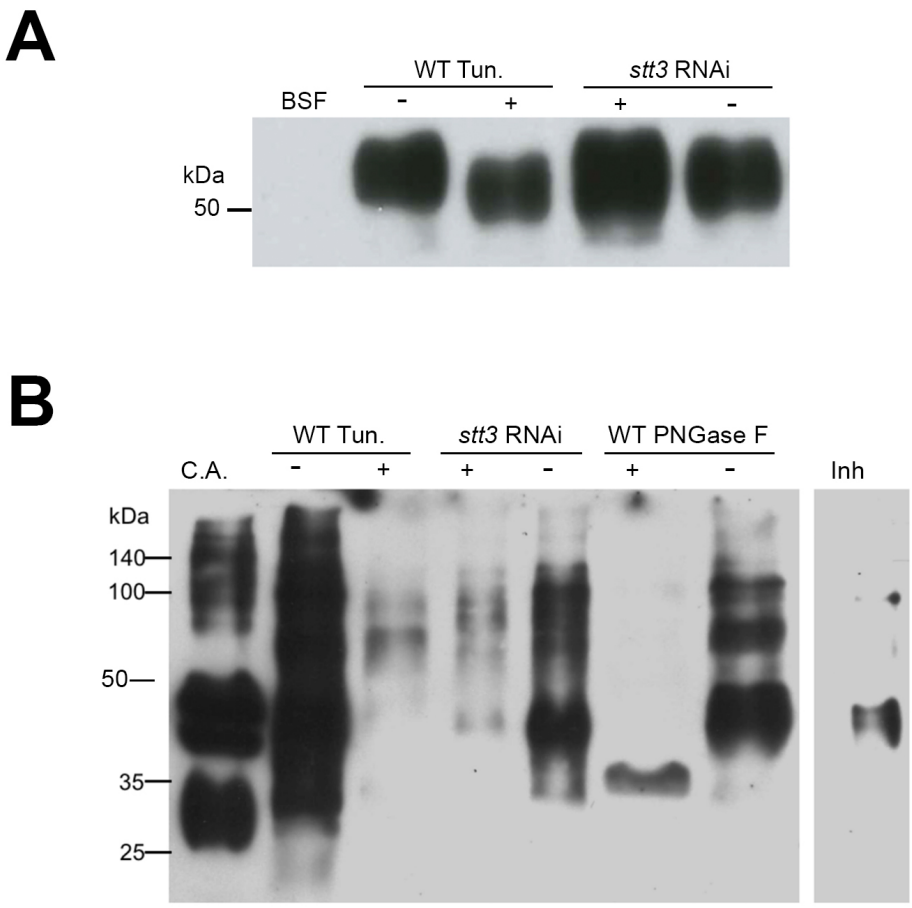


Figure 5.5. Characterisation of procyclins from s*tt3* RNAi cells. (A) Western blotting using anti-EP procyclin on extracted procyclins from wild type untreated (Tun. -) or tunicamycin-treated (Tun. +) cells, induced (+) or uninduced (-) RNAi cells, and cell lysate from BSF as negative control. **(B)** ConA blotting of extracted procyclins from wild type tunicamycin-treated or untreated cells, induced or uninduced RNAi cells. Commercial, glycosylated chicken albumin (C.A.) was used as positive control. Specificity controls: wild type procyclins either treated or not with PNGase F, pre-inhibited ConA on wild type procyclins (Inh).

5.2.2.3 *stt3* RNAi cells show a partial infection phenotype in tsetse flies

The *stt3* RNAi cell line has shown significant reduction in *N*-glycosylation activity. To study the role(s) and essentiality of *N*-glycosylation regarding infectivity and development in the tsetse, the *stt3* RNAi cell line was pre-induced in culture and

used to infect tsetse to score MG, PV and SG infections. The rate of infection in non-induced were 92% MG infection, 40% PV infection and 12% SG, which is comparable to that of wild type cells and indicated no “leakiness” in the inducible system (**Figure 5.6**). This is in clear contrast to the results observed in induced cells, which showed an overall reduction in infectivity (i.e. ~15% and ~62% in MG and PV, respectively, and no parasites detected in SGs) (**Figure 5.6**).

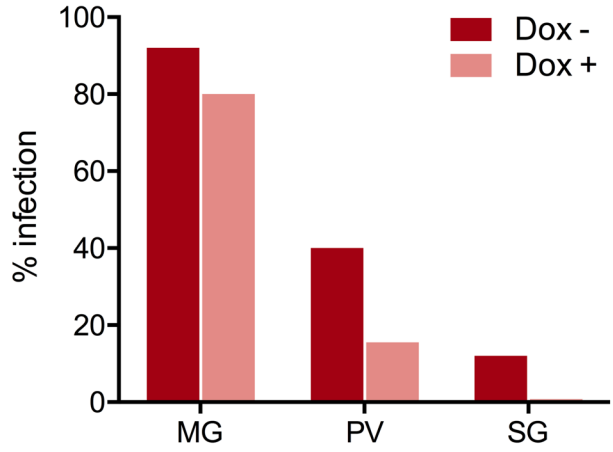


Figure 5.6. Essentiality of *N*-glycosylation for tsetse infection. Percentage of infection in MG, PV and SG at 30 d.p.i. of *stt3* RNAi cells either uninduced (Dox -) or induced (Dox +).

5.2.3 Systematic collection of *T. brucei* metacyclic cells for glycan structural analysis

The *N*-glycans in the MCF glycocalyx are expected to be predominantly attached to mVSGs as in BSF VSGs. After the findings in Chapter 3, however, more *N*-glycans could be expected to be found in MISP and BARP. Despite VSGs and mVSGs are assumed to be structurally equivalent, their respective attached *N*-glycans may not have conserved structures as these depend in part on the STT3 isoforms functioning in each stage of the parasite (Izquierdo et al., 2009b). As shown in section 5.2.1, BSF predominantly express *stt3-B* and *stt3-C* but also *stt3-A.* However, in SG parasites (mainly MCF) only expression of *stt3-B* could be detected, suggesting that the *N*-glycan structures in mVSGs may not be the same as in BSF.

To determine the structure(s) of MCF surface *N*-glycans, multiple tsetse infections were set up using either BSF from mouse infected blood, PCF or cultured BSF (strain AnTat 1.1 90:13). Parasites infecting the SG were collected at 30 days post infection. MCF (~95% of SG forms) were isolated from the SGs, washed in PBS to remove salivary proteins and snap-frozen until further use. The equivalent to ~250 SG-infected flies (35 infection experiments, ~3,500 total flies) were pooled together to further perform a glycoproteomics analysis in order to determine the structure of the *N*-glycans and the proteins these belonged to. Although the sensitivity of these methods has greatly improved over the recent years, the required number of cells for these approaches is normally ~10^8-10^9. The estimated number of collected metacyclics during this thesis is ~$2x10^6$, considering an average number of $8x10^3$ MCF cells per infected fly. Consequently, due to time constrains and the impossibility to obtain the desired number of metacyclics for mass spectrometry analysis, the glycoproteomics analysis of *T. brucei* metacyclic, extracted from infected tsetse SGs, was never performed. However, an interesting observation can be made after performing the tens of tsetse infection for MCFs collection during a period of ~3 years. It became clear that flies infected with BSF from infected blood yielded higher SG infection rates (~27%) compared to infections carried out using either PCF (7.6%) or cultured BSF (0%) (**Figure 5.7**).

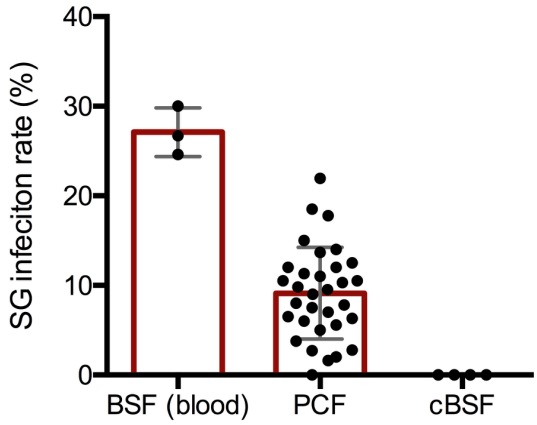


Figure 5.7. SG infection rates obtained in multiple tsetse infection for MCF collection. Comparison of the SG infection rates between tsetse infected with either BSF from mouse blood, PCF or cultured BSF (cBSF). Bar notes the mean; error bars show standard deviation. Each dot corresponds to an infection event of at least 100 flies.

5.3 Discussion

The glycobiology of *T. brucei* has been widely studied over the years, mostly focus on BSF from animals and PCF due to the high number of cells that can be obtained from either stage. Besides structural studies which determined the major *N*-glycan structures found on *T. brucei* (Ferguson, 1997), the study of the *N*-glycosylation and GPI anchor pathways in this organism has provided a massive contribution to the entire field of glycobiology (Ferguson et al., 1988). However, little is known about the *N*-glycan structures found in advanced stages of the parasite in the tsetse (i.e. EMF and MCF) and the role of *N*-glycans regarding parasite development in the vector. This chapter provided more insights into the essentiality of parasite *N*-glycosylation in the vector.

5.3.1 Developmental regulation of the trypanosomal STT3 oligosaccharyltransferase isoforms

Understanding how the three *stt3* paralogs are expressed throughout the life cycle of *T. brucei* is key to better understand the function and structure of parasite *N*-glycans in response to adaption to the different environments. In particular, with the exception of procyclic cells, very little is known about the type of glycosylation present in most parasite stages developing within the tsetse. Here, the RNA expression of *stt3-A, stt3-B* and *stt3-C* was determined in all *T. brucei* stages. It was confirmed that BSF express both *stt3-A* and *stt3-B* isoform at similar rates while, in comparison, PCF mainly express *stt3-B*. This goes in agreement with a previous study on *stt3* expression (Izquierdo et al., 2009b) and with the reported *N*-glycan structures predominantly found in BSF (complex type as a result of *stt3-A* activity) or in PCF (high-mannose type as a result of *stt3-B* activity). While previous reports failed to detect any expression of *stt3-C* in neither BSF nor PCF, in this chapter I showed evidence of a significant expression of *stt3-C* in BSF, PCF and MG procyclics. This could be related to the use of strain AnTat 1.1, which is known to be pleomorphic and fly-transmissible. In contrast, previous studies used a transgenic clone of strain Lister 427, which has been highly passaged in culture and is non-pleomorphic and not fly-transmissible. Because it is unknown how STT3-C functions and considering the high sequence identity between STT3-A, STT3-B and STT3-C (~88%), it is possible that expression of STT3-C, in the Lister 427 strain, may have been repressed as a result

of culture adaptation and increased fitness. It was hypothesised that, because *stt3-C* was absent in both BSF and PCF, *stt3-C* could play its role in the tsetse stages, specifically in EMF since these are the most divergent to PCF and BSF. It could be also hypothesised that MCF would have similar expression patterns to BSF, as these cells re-express the VSG coat ready to be transmitted to the mammalian host (Tetley et al., 1987). This does not seem to be the case considering these results as it was found that MG procyclics, PV and SG stages predominantly express *stt3-B*. It is important to consider that the STT3 activity also depends on the residues flanking the *N*-glycosylation sequons (Izquierdo et al., 2009b; Jinnelov et al., 2017) and so the peptide sequence may influence the *N*-glycan structure output. STT3 specificity has been also determined in other parasites such as *Leishmania* ((Hese et al., 2009; Nasab et al., 2008) and *Plasmodium* (Samuelson et al., 2005), the latter transferring extremely truncated glycan precursors (GlcNAc only or a chitobiose core). The biological significance of this broad spectrum of OST subunits remains unclear.

Analyzing gene expression in tsetse stages is always challenging mainly due to two factors. First, the limited number of cells yields very low quantities of RNA, and second, the heterogeneity of parasite stages and tsetse cell debris in these samples may interfere with the results. It cannot be assumed that *stt3-B* is the only paralog expressed in the tsetse stages but the most abundant one. Some cell types, especially EMF in the PV (which represent 5-10% of the total PV population) or attached EMF in the SG (variable prevalence in SG, co-habiting with MCF), may have different expression patterns of *stt3* but these could be masked by the predominant cell types.

Lastly, it is interesting to notice that BSF have the overall highest expression of *stt3,* and the expression levels go down in the tsetse. This increased need for *stt3* expression in BSF could be attributed to the huge demand of *N*-glycans for VSGs. BSF keep a massive rate of continuous and fast production of VSGs, which indirectly implies that the *N*-glycosylation pathway needs to increase to meet the need.

5.3.2 Functional studies on *N*-glycosylation in the tsetse stages

Attempts for generating conditional-null trypanosome mutants, defective in *N*-glycosylation, were unsuccessful for two different genes. The genes of choice were the *stt3s* (*stt3-H, stt3-A, stt3-B, stt3-C*) and *rft1* as they encode key enzymes in the

N-glycosylation pathway. The STT3 oligosaccharyltransferase transfers the *N*-glycan precursor to the protein, and the putative flippase RFT1 flips in the $Man_5GlcNAc_2$-PP-dol precursor glycan from the cytosolic side of the ER into the luminal face during glycan assembly (Helenius et al., 2002). Remarkably, a double knockout of *rft1* has been generated in *T. brucei* PCF (Jelk et al., 2013). In the case of *rft1*, both were knocked out but surprisingly, *N*-glycans were still being produced although with altered structures and in lower quantities (Jelk et al., 2013), suggesting that there may be an alternative flippase activity form either an unknown enzyme or by spontaneous diffusion through the ER membrane. On the other hand, trypanosomal *stt3* genes could never be completely removed, being a single knockout enhanced with *stt3* RNAi the best approach developed to deplete global *N*-glycosylation (Izquierdo et al., 2009b). All these mutants were created in the Lister 427 strain (which is non-fly transmissible) making the tools unsuitable for studies in the tsetse. In this thesis, another attempt of generating *stt3* and *rft1* conditional-null mutants in AnTat 1.1 90:13 was carried out. After multiple unsuccessful attempts, clonal cell lines were obtained resistant to the knockout antibiotic markers for both *stt3* and *rft1*. A PCR-based strategy was designed to screen and genotype these cell lines for the identification of the insertion of resistance markers, addbacks and removal of the endogenous genes. Results from several cell lines showed that, despite having successfully inserted the two resistance markers and removed both alleles from the endogenous loci, all clones revealed the presence of at least one copy of the endogenous genes. Because a ~96% of the total *T. brucei* proteins are *N*-glycosylated, it can be hypothesised that during the transfection processes, gene duplication events may have taken place as a result of applying pressure to deplete such important genes. This is not unprecedented as *T. brucei* glycosylation mutants have been described to recombine some of the *stt3* isoforms after selection with cytotoxic carbohydrate binding molecules (Castillo-Acosta et al., 2016; Castillo-Acosta et al., 2013). Further analyses based on southern blotting could confirm this hypothesis and provide more information on when these events happened. In future attempts, an alternative strategy could consist on inserting the addback gene in first place and keep its expression active by default, so the enzymatic activity does not cease during deletion of the endogenous alleles, which may prevent gene duplication.

Having failed in creating conditional-null mutants, a second approach based on gene silencing through RNAi was considered. *Stt3* silencing was preferred over *rft1* since this proved to be effective in reducing *N*-glycosylation levels (Izquierdo et

al., 2009b) but not altering glycan structures (Jelk et al., 2013). The selected RNAi cell line showed acceptable downregulation of *stt3* RNA, which indeed led to a reduction of EP procyclins glycosylation as proven by FACS and lectin blotting analyses, although not as effective as the reduction observed in tunicamycin-treated parasites. RNAi cells were then used to infect tsetse to study the impact of decreased *N*-glycosylation in parasite infectivity. Tunicamycin-treated cells were not considered in this experiment since the drug becomes lethal after a few days. Remarkably, a reduced overall *N*-glycosylation by the *stt3* knockdown had a negative impact on parasite infectivity of the tsetse in MG, PV and SG. This is in contrast to published results showing that alteration of the parasite *N*-glycosylation after deletion of *Alg3* gene, does not affect infectivity of the tsetse MG (Manthri et al., 2008). The migration and infection of these tissues are naturally major bottlenecks for *T. brucei* transmission. The parasite is always more effective at infecting the MG (~90%) and infectivity rates decrease along with development in the PV (~40%) and SG (~10%). Therefore, the effect in infectivity due to a reduction of *N*-glycosylation is little in the MG and becomes greater in the PV and SG. Strikingly, the *stt3* RNAi led to a complete absence of SG infections compared to the control.

As to why *N*-glycosylation may be important for parasite progression through the fly is unknown, but one straightforward explanation is that surface *N*-glycans contribute to shield the parasite surface against the action of MG hydrolases (Acosta-Serrano et al., 2001). In addition, recent evidence suggest that glycosylation mutants are unable to perform social motility (SoMo) *in vitro*, although whether SoMo also occurs in the tsetse remain to be proven (Imhof et al., 2015).

CHAPTER 6

New methods for imaging African trypanosomes in the tsetse

6.1 Introduction

T. brucei has been extensively studied and characterised *in vitro.* However, studies on vector-parasite interactions are scarce and, remarkably, one major bottleneck for these studies is the low quality of parasite imaging inside tsetse tissues.

Transmission and scanning electron microscopy were used many years ago in the firsts attempts to visualise *T. brucei* in the salivary glands of the tsetse (Gluenz et al., 2015; Sharma et al., 2008; Tetley et al., 1987; Tetley and Vickerman, 1985). On the other hand, light sheet microscopy was recently used to facilitate imaging of trypanosomes in the tsetse MG and PV (Schuster et al., 2017). However, access to this type of state-of-art microscope is difficult, including the possibility of imaging live or fixed pathogens in dedicated Cat 3 facilities. If higher magnification and detail is required, trypanosomes can be isolated from the tsetse, stained and imaged by regular fluorescence microscopy methods, but consequently the biological context is lost. In this chapter, I describe imaging methodologies based on confocal laser scanning microscopy. A series of staining and mounting methods were developed to allow the observation of either naïve or trypanosome-infected tissues from tsetse flies in their native context.

Two examples of applicability of these methodologies are reported. First, I describe how the 3D reconstruction of several tsetse organs was used to study the tissue damage caused upon exposure to nitisinone, a powerful inhibitor of the tyrosine metabolism that kills all blood feeder insect vectors, including tsetse (Sterkel et al., 2016) and (in preparation). Second, I greatly exploited these methodologies to study a novel route of colonisation of the tsetse gut by *T. brucei*.

6.2 Methods

6.2.1 Tsetse nitisinone feeding and tissue processing

Teneral (<24h old) male tsetse were blood-fed on sterile defibrinated horse blood (TCS Biosciences) using standard silicon membrane feeding protocols. Three days later, flies were offered a second meal supplemented with 10 μg/mL wheat germ agglutinin (WGA)-rhodamine (Vector Laboratories) in either normal horse serum or IPL41 serum-free insect medium (Sigma), with or without 500 ng/mL (2-[2-nitro-4-(trifluoromethyl)benzoyl]cyclohexane-1,3-dione (nitisinone) (Sigma). Tsetse were anaesthetised on ice 15 hours post feeding and the following tissues were dissected in ice-cold PBS: midgut (with attached proventriculus), flight muscle, fat bodies, Malpighian tubules and salivary glands. Tsetse survival was determined 72 hours post nitisinone feeding. Dissected tissues were immediately fixed in fresh 4% PFA for 1 hour at RT, washed and stained with SiR-actin (1:1000 dilution in PBS, Cytoskeleton Inc.) for 3 hours at RT. Tissues were posteriorly incubated in 500 ng/mL DAPI for 10 minutes at RT, washed and mounted in 1% low-melting agarose at <40°C mixed with Slowfade Diamond mounting oil in a contained (higher volume) area on a glass slide. Slides were imaged using a Zeiss LSM-880 and tissues reconstructed from a series of z-stacks as described in section 2.5.3. SiR has excitation and emission peaks at 652 nm and 674 nm, respectively.

6.2.2 Live staining of the tsetse peritrophic matrix

To stain the peritrophic matrix along the midgut, flies were fed every day with FBS supplemented with 15 μg/mL WGA-rhodamine. WGA is a lectin that binds to terminal linear GlcNAc residues. As proteins (peritrophins) in the tsetse PM are crosslinked with chitin fibres, WGA was used as a probe for staining of this tissue. This method is compatible with imaging of fluorescent (GFP) trypanosomes infecting the MG and PV. Alternatively, to stain the PM at the synthesis site, flies were fed with FBS 15 μg/mL WGA-rhodamine for 10 minutes and the PV, still attached to the anterior midgut, was dissected out after 15 minutes post-feeding. Exceeding this time results in a loss of staining in the PV since the rate of PM synthesis and serum meal pumping are particularly fast (Lehane and Msangi, 1991). This method is also compatible with imaging of fluorescent trypanosomes infecting the PV and anterior

midgut. However, imaging of intact midgut is challenging because of the damage occurring due to the expansion caused by the blood/serum meal.

6.2.3 3D reconstruction of trypanosome-infected tsetse proventriculus

To study the early infections of the tsetse PV by *T. brucei* parasites, tsetse control or infected with either *T. brucei* J-10 (GFP expressor) or AnTat 1.1. 90:13 (mNeonGreen; this thesis) were fed every day on serum meals composed of FBS supplemented with 15 μg/mL WGA-rhodamine for PM staining as described in section 6.2.2. At 5 dpi, the PV was immediately dissected after feeding and fixed in fresh 1% PFA at 0°C for 1 hour to preserve the trypanosome fluorescent protein fluorescence. Tissues were then stained with SiR-actin (1:1,000 dilution) for 4 hours on ice, stained with DAPI 300 ng/mL for 10 minutes, and mounted in 1% low-melting agarose (<40°C) mixed with Slowfade Diamond mounting oil, in a higher-volume contained area on a glass slide. Specimens were imaged and reconstructed using a Zeiss LSM-880 as described in section 2.5.3.

6.2.4 *Ex vivo* 3D reconstruction of trypanosome-infected peritrophic matrices

Tsetse infected with *T. brucei* J-10 strain (GFP) were dissected at 9 days post infection to isolate the PM in PBS. The PM was washed several times in PBS and allowed to attach to poly-L-lysine slides. The PM was fixed in 1% PFA at 0°C for 1 hour, washed and stained with 10 μg/mL WGA-rhodamine and 300 ng/mL DAPI for 15 minutes. After washing, slides were mounted in Slowfade Diamond mounting oil and imaged and reconstructed using a Zeiss LSM-880 as described in section 2.5.3.

6.3 Results

6.3.1 3D reconstruction of tsetse organs to investigate tissue damage upon nitisinone exposure.

The methods described in section 6.2.1 were applied to investigate the tissue damage caused after exposure to nitisinone (also referred as NTBC). Several groups of tsetse were fed with a fixed sub-lethal dose of NTBC (500 ng/mL) dissolved in either horse serum or IPL41 serum-free insect medium, supplemented with WGA-rhodamine lectin for PM staining. Negative controls consisted only of horse serum or IPL41 medium. In previous experiments (unpublished data), the meal medium composition was shown to modulate toxicity of NTBC, the IPL41 medium being more effective in killing compared to serum using identical NTBC doses. It was therefore hypothesised that they would lead to “early” and “late” tissue damage phenotypes, respectively. At 15 hours post feeding, flies were dissected to collect the following organs (**Figure 6.1-A**): midgut (with attached proventriculus), flight muscle, Malpighian tubules, fat bodies and salivary glands. All of them were shown to be severely damaged upon NTBC exposure in previous experiments. In addition, flight muscle was of particular interest since paralysis is the first visible effect of NTBC, and flies show muscle detachment.

Tsetse survival was determined 72 hours post-feeding to relate degree of tissue damage with death rates (**Figure 6.1-B**). The NTBC (serum) “early phenotype” group showed a survival rate of ~34% and thus proved to be a sub-lethal bloodmeal compared to the 0% survival of the NTBC (IPL41) “late phenotype” group. Both control groups had higher survival rates in comparison (92% serum, 52% IPL41), although the IPL41 itself seems to have an intrinsic toxic effect in some flies. However, the aspect of dead flies due to NTBC is clearly distinct from dead flies in the control groups. Control and affected tissues were then processed as described in section 6.2.1 and imaged by LCSM for analysis. The tissues were reconstructed as a 3D overview model generated from several z-stacks combining DAPI and SiR-actin stainings. Detailed sections of the tissues were also imaged to analyse damage at cellular level.

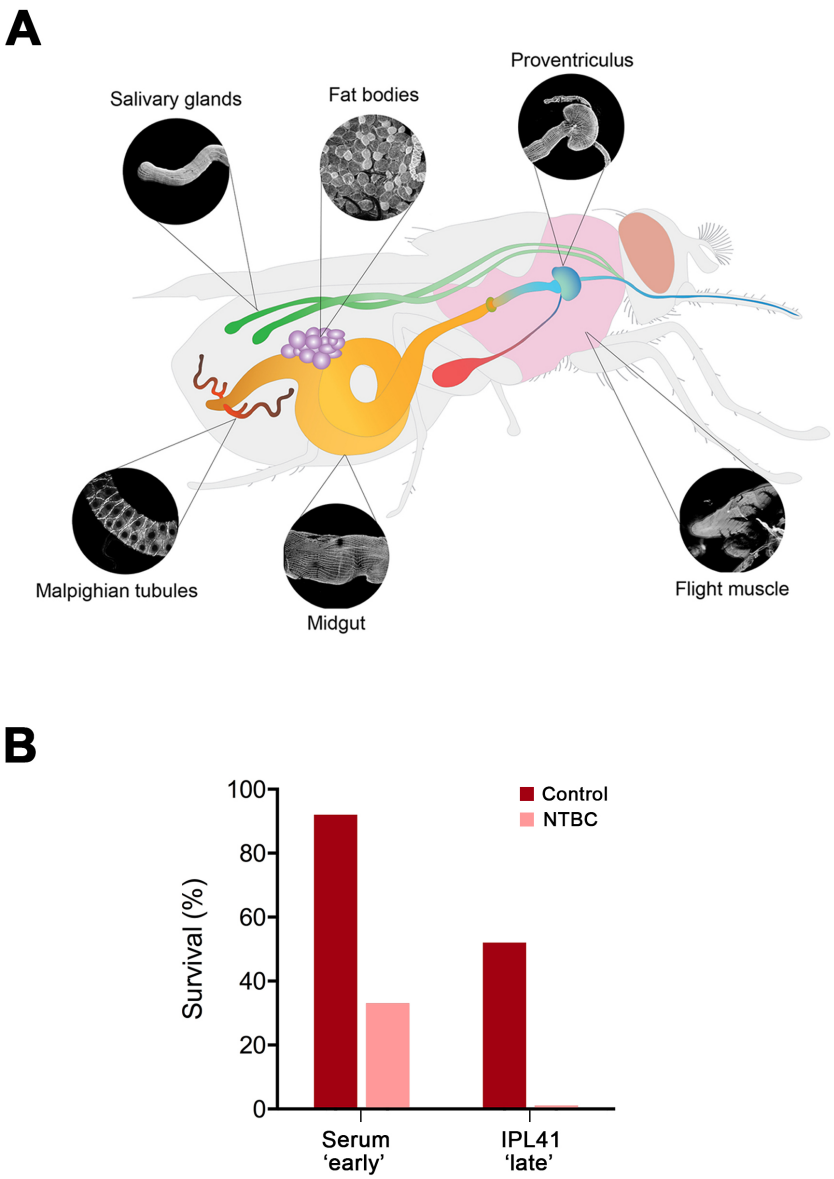


Figure 6.1. Tsetse tissues investigated for damage upon NTBC exposure. (A) Schematic of a tsetse fly highlighting the organs dissected out for imaging. **(B)** Survival rates (percentage) at 72 hours post NTBC ingestion.

After reconstruction, the tsetse anterior midgut (AM) was shown to be formed by an outer layer of muscle cells which are organised in longitudinal fibres (along the midgut length) and transversal fibres (around the midgut perimeter). The muscle fibres surround an inner layer made up by rounded epithelial cells. The control AMs could be dissected out intact and connected to the PV, crop and oesophagus, while the NTBC-treated tissues (early phenotype) appeared to be structurally damaged and

the AM disconnected from PV, crop and oesophagus (**Figure 6.2**). When observing the AM muscle fibres at higher magnification it could be observed that, while the longitudinal muscle fibres did not appear to be altered upon NTBC treatment, the transversal fibres were significantly fewer, thicker and less compact compared with the controls. It is reasonable to think that such alterations in tissue organisation make these important organs unfunctional, eventually leading to death. The same organs could not be isolated from NTBC-treated flies with "late phenotype": at this stage, most organs have dissolved forming a single tissue mass in the abdomen which, after reconstruction, was shown to have unstructured remains of midgut tissue, Malpighian tubules and fat bodies (**Figure 6.2**).

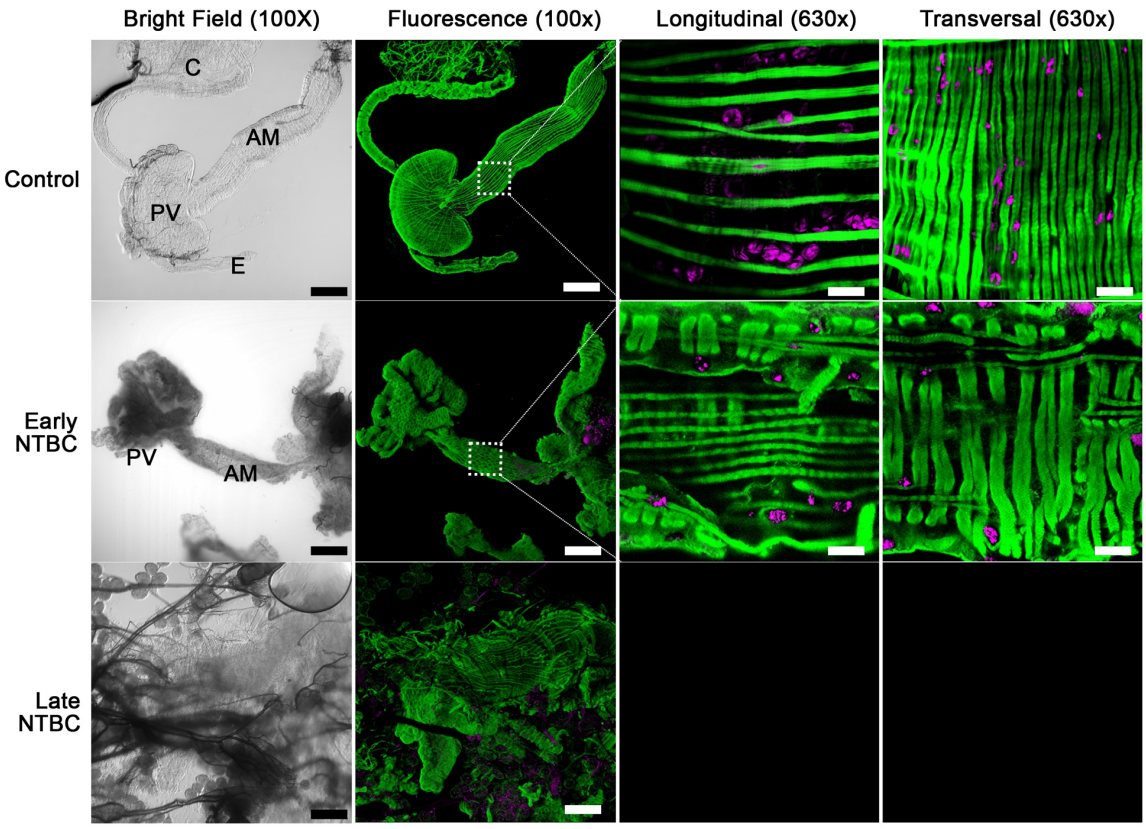


Figure 6.2. Midgut and proventriculus reconstructions. Bright field and 3D fluorescence reconstruction overviews (100x) of tsetse anterior midgut (AM), proventriculus, crop (C) and oesophagus (E) from control, NTBC-fed "early" and "late" phenotype flies. Fluorescence merges SiR-actin staining (green) and DAPI (magenta). Detail of either horizontal or vertical muscular fibres (630X) from the anterior midgut (AM). Scale bar = 200 μm (100X) and 20 μm (630X).

Also, the intensity of the SiR-actin staining was >10-fold less than the control and “early phenotype”, suggesting there has been an overall loss of actin due to cell death, which made it impossible to image muscle fibres in detail (also from flight muscle). In addition, a systemic darkening of the tissues, as a result of a systemic myelinisation process, was observed in NTBC-treated flies.

The flight muscle was reconstructed to compare the architecture of the fibres and sarcomeres in detail between samples in equivalent areas of the muscle (**Figure 6.3**). The overall structure of the muscles was not found to be altered in NTBC-treated flies. The “late phenotype” flies showed signs of muscle detachment, but the overall integrity of the fibres was maintained. When looking in detail at the level of single muscle fibres, the untreated samples showed well organised and straight fibres, with clearly defined sarcomeres. In ‘early phenotype’ flies, muscle fibres were still visible although in many areas the gaps between fibres and sarcomeres had increased in size, some fibres lost linearity and the nuclei became smaller. Samples from “late phenotype” flies showed a dramatic decrease in actin content, the fibres were more compacted, significantly thinner and the sarcomere unit was lost, suggesting these muscles completely collapsed and became unfunctional.

The fat bodies, Malpighian tubules and salivary glands were also analysed to see the damage extent beyond the midgut and muscle. In previous experiments, NTBC was shown to make the Malpighian tubules collapse, preventing the tsetse from excreting NTBC in first instance. The fat bodies are known to be involved in detoxification (Arrese and Soulages, 2010) and thus were also taken in consideration for analyses. Lastly, although not directly involved in digestive processes, salivary glands were also analysed since they share the abdominal space with the rest of organs mentioned above. The fat bodies were slightly affected by NTBC exposure (**Figure 6.4**). While in untreated flies they can be isolated in mass containing hundreds of them, “early” phenotype flies contained less and smaller fat bodies, with a slenderer shape. Although oenocytes could still be found, SiR-actin staining in all cell types was also less abundance of actin. In “late” phenotype flies, remains of fat bodies were still recognisable but scarce and isolated within the mass of abdominal tissues, and oenocytes could not be found. The poor staining of SiR and DAPI showed areas of accumulation within the bodies, likely as a result of cell death processes.

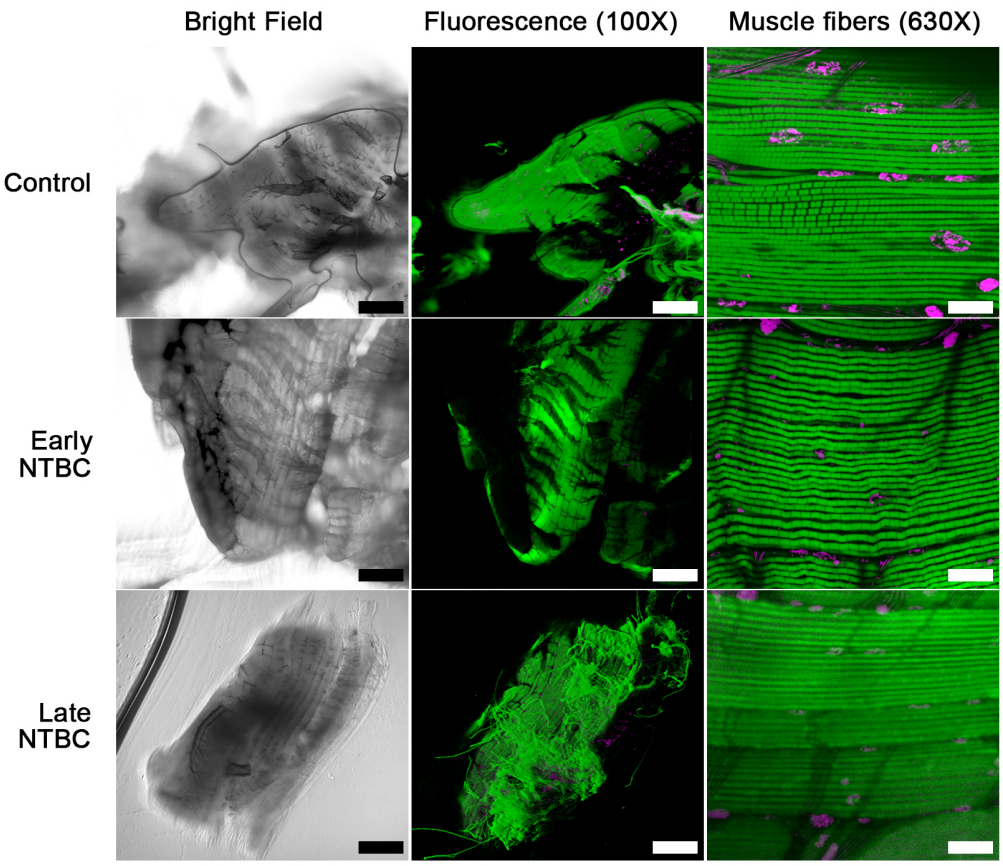


Figure 6.3. Reconstruction of the tsetse flight muscle. Overview (100X), and detail of the muscle fibres (630X) from control and NTBC-treated "early" and "late" phenotype flies, in bright field and fluorescence reconstructions (SiR-actin in green, DAPI in magenta). Scale bars = 200 μm (100X) and 20 μm (630X).

The Malpighian tubules were reconstructed to observe that the overall integrity is not compromised in "early" phenotype flies despite being unable to excrete digestion products. However, as seen in the other tissues, actin was greatly lost. Malpighian tubules in "late" phenotype flies could not be found (**Figure 6.5**).

Lastly, the salivary glands were not found to be especially affected in NTBC "early" phenotype samples as they kept their integrity and the muscle fibres only seemed to be less tight compared with the control (**Figure 6.6**). Remarkably, in "late" phenotype flies salivary glands could not be found as they were probably dissolved or lost their integrity.

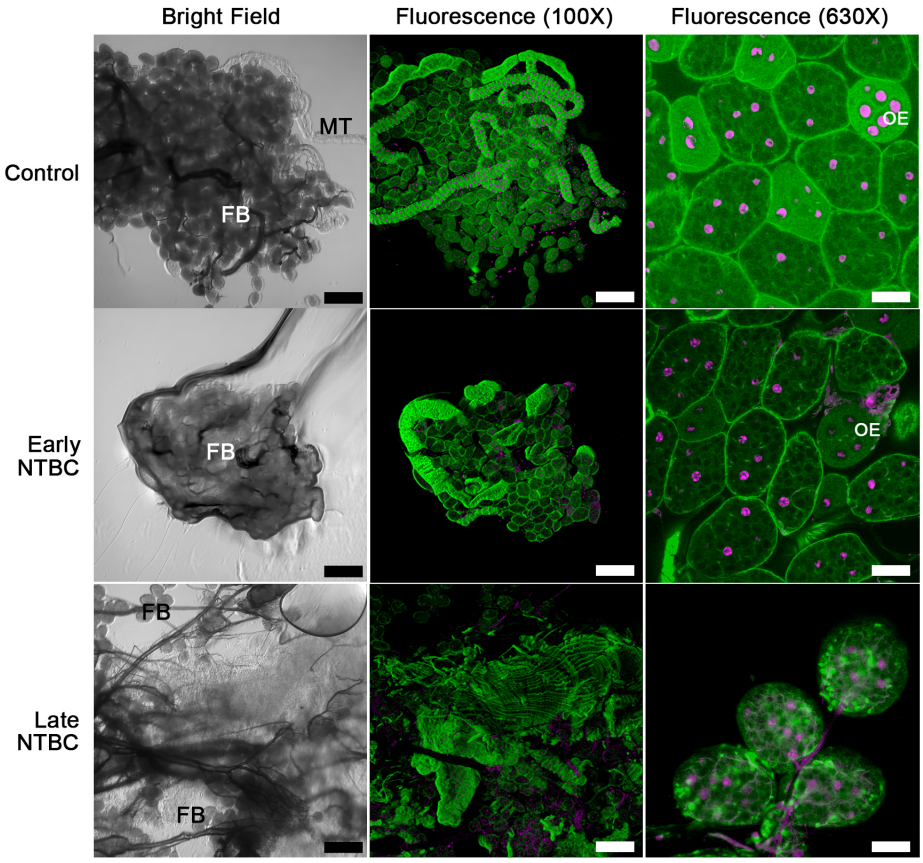


Figure 6.4. Tsetse fat bodies reconstructions. Overview (100X) and close detail (630X) of fat bodies from control and NTBC-treated "early" and "late" phenotype flies as bright field and fluorescence reconstructions (SiR-actin in green, DAPI in magenta). Fat bodies mass (FB) may contain Malpighian tubules (MT). Scale bars = 200 μm (100X) and 20 μm (630X).

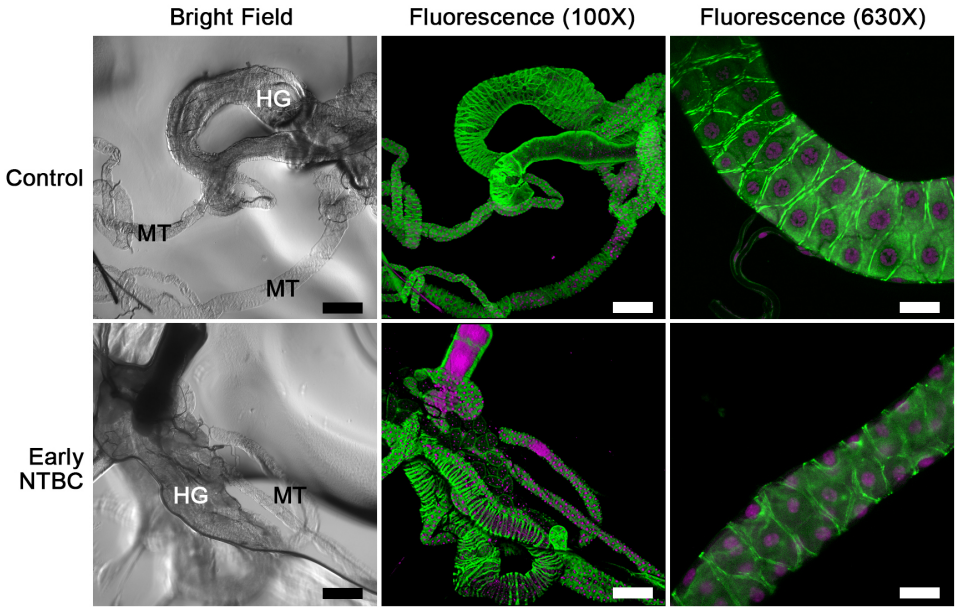


Figure 6.5. Malpighian tubules reconstructions. Overview (100X) and close detail (630X) of samples of Malpighian tubules from control and NTBC-treated "early" flies as bright field and fluorescence reconstructions (SiR-actin in green, DAPI in magenta). Malpighian tubules (MT) and hindgut (HG) noted in bright field. Scale bars = 200 μm (100X) and 20 μm (630X).

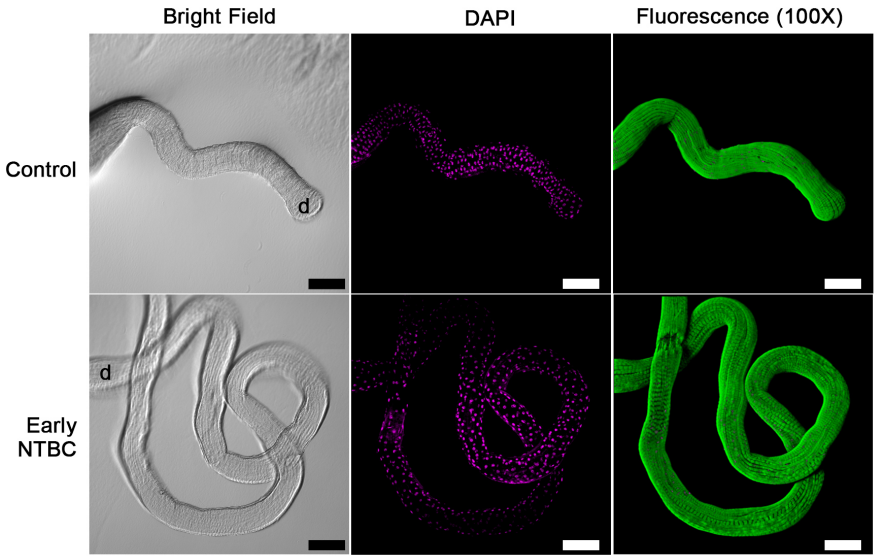


Figure 6.6. Tsetse salivary gland reconstructions. SG from control and "early" phenotype NTBC in bright field, DAPI and 3D fluorescence using SiR-actin (green) and DAPI (magenta). SG distal end pointed as 'd' in bright field. Scale bar = 100 μm.

6.3.2. Studies on *T. brucei*-tsetse interactions using imaging methods

6.3.2.1 An overlooked *T. brucei* early infection of the tsetse proventriculus

It is well accepted that when the tsetse uptakes a *T. brucei*-infected bloodmeal, the parasite first colonises the MG and after 1-2 weeks it migrates towards the PV (Gibson and Bailey, 2003). However, here we observe that flies fed with infected rat blood contained a few parasites in the PV as soon as 2 d.p.i and developed a regular heavy PV infection from 3 dpi (**Figure 6.7-A**). Time course experiments showed that, while the infection in the MG remains stable during the first 5 days post infection, PV infection rates rapidly increase from ~8% (2 d.p.i) to ~77% (5 d.p.i) (**Figure 6.7-B**). Interestingly, this capability of PV early infection seems to be lost upon parasite adaptation to culture. BSF blood parasites were put in culture and maintained for 9 days before fed to the tsetse. These recently-adapted BSF showed a dramatic decrease in PV infection rates (37%) at 5 d.p.i compared to BSF blood parasites (78%), and the early PV infection was totally lost in a BSF cultured cell line (0%) (**Figure 6.7-C**). PV infection rates at 15 d.p.i, however, are high and similar in all groups. This suggests that culture adaptation negatively affects PV invasion only at the early stages. It remains to be investigated what is the role of the early PV infection regarding parasite development or transmission. It can be thought that the early PV infection is somehow linked to the transmission index. Infection rates in SG were assessed in flies infected with the previous parasite groups. Blood BSF lead to the highest SG infection rates (22%) and these decreased in parasites recently-adapted to culture for 9 days (8.7%). Cultured BSF are completely unable to develop a SG infection (0%) (**Figure 6.7-D**). Lastly, MG and PV parasites at 5 d.p.i and 15 d.p.i were extracted for stage identification based on morphology, nucleus/kinetoplast relative position and cell length (**Figure 6.7-E**). Early PV parasites were found to be 100% mesocyclics with same morphology, DNA relative position and cell length than those found at 15 d.p.i. However, early infected PVs do not contain EMF forms while at 15 d.p.i. they contain both long and short EMFs. This suggests that parasites infecting the PV at early stages are indeed proventricular forms and not procyclics that came in from the MG, so the development and invasion process could be equivalent to that at 15 d.p.i. with the exception of EMFs which are not yet formed.

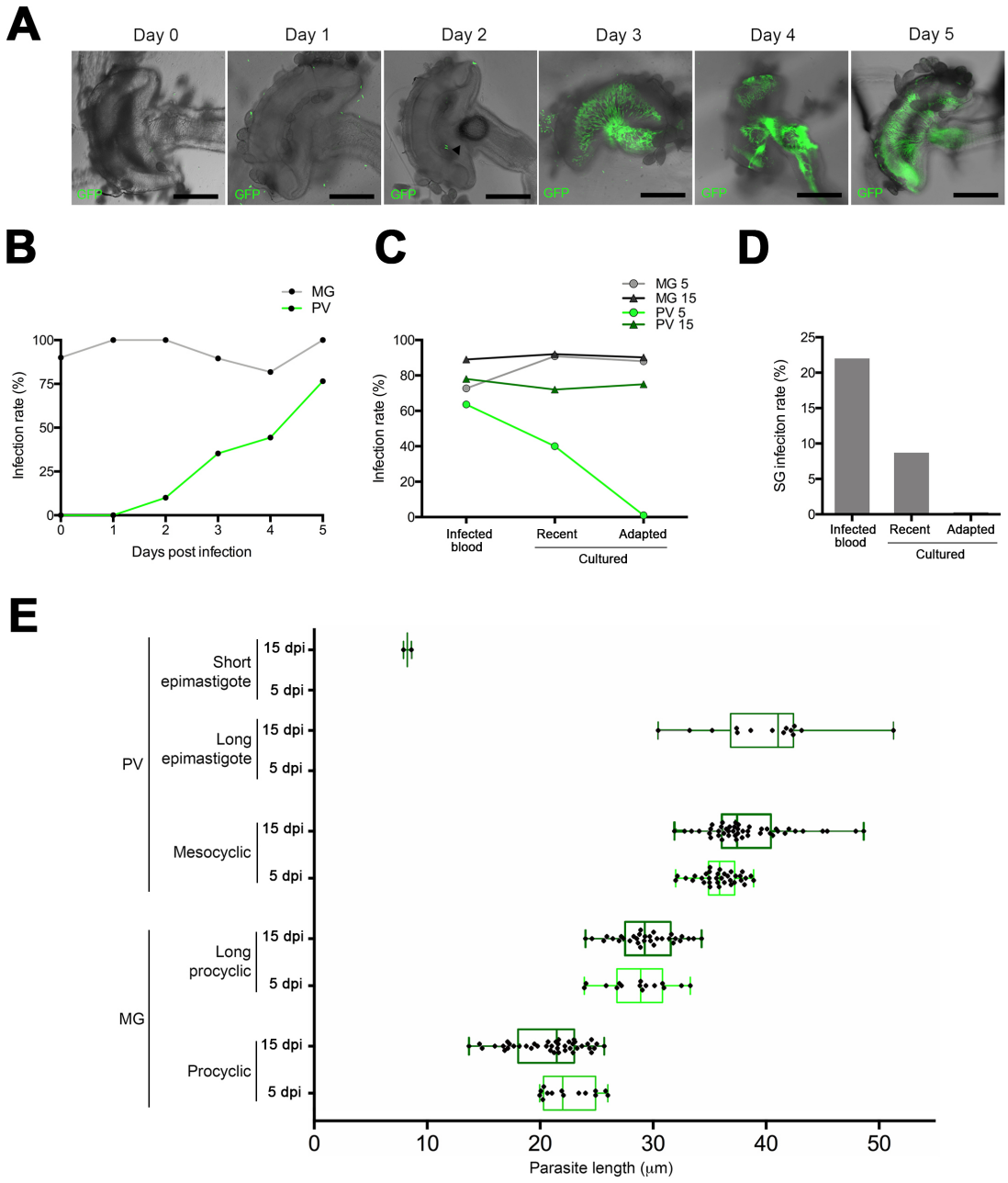


Figure 6.7. *T. brucei* early infection of the tsetse proventriculus. (A) PV infection time course (0 to 5 d.p.i) in flies infected with *T. b. brucei* (strain J-10, constitutive GFP); scale bars = 200 μm. **(B)** Infection rates from time course in 'A', in MG and PV. **(C)** Infection rates (5 and 15 d.p.i) of MG and PV from tsetse infected with BSF (strain AnTat 1.1. 90:13) either from mouse infected blood (Blood), adapted to culture for 9 days (Adapted) or cultured cells (cBSF). **(D)** Salivary gland infection rates of flies infected with the same groups in 'C' at 30 dpi. (**E**) Cell length of parasites stages found in the MG and PV at 5 and 15 d.p.i.

6.3.2.2 Imaging the early-infected tsetse proventriculus

It is also well established in the community that *T. brucei*, after colonizing the MG, needs to reach the ectoperitrophic space (ES; space between the peritrophic matrix and the MG endothelial cells) (Gibson and Bailey, 2003). It is assumed that the parasites cross the PM to do so, although there no direct evidence has been reported. Once in the ES, it is assumed that procyclic trypanosomes migrate towards the PV where the PM is synthesised. The discovery of the PV early infection challenges this dogma, since the time window is too short for the parasites to carry out the crossing and migration processes, and potentially suggests that reaching the ES could be a consequence of this PV early invasion.

The early-infected proventriculus was stained with different markers to better understand the invasion process and interactions between parasites and tsetse tissue. The overall PV actin-based reconstruction was made using the same labelling and mounting method as in section 6.2.1 plus staining of the PM using the lectin WGA, DAPI to stain nuclei and GFP fluorescence from parasites. By analysing several z-stack sections, orthogonal views were generated and used to determine the precise localisation of the parasites in the PV and AM. In **Figure 6.8-A** it is depicted the areas of interest using a control and infected PV as example. Thanks to the different staining tools, it is clear to see the limits of the PV, the AM connected to it, the endothelial cells and the ES. A control and infected PV orthogonal views were analysed in detail (**Figure 6.8-B-C**). Parasites can be found in the PM lumen but also in the ES of the AM where they seem to gather in small areas along the ES.

Sequentially added fluorescence channels in 3D reconstructed control and early-infected PV, at angles of 20° and 45°, help to understand how the PM is secreted inside the PV from multiple areas becoming a cylinder-shaped tube that goes along the MG. It also shows how the parasites invade the PV outside the PM (by the ES) but underneath the main section of PV columnar and muscular cells (**Figure 6.9**).

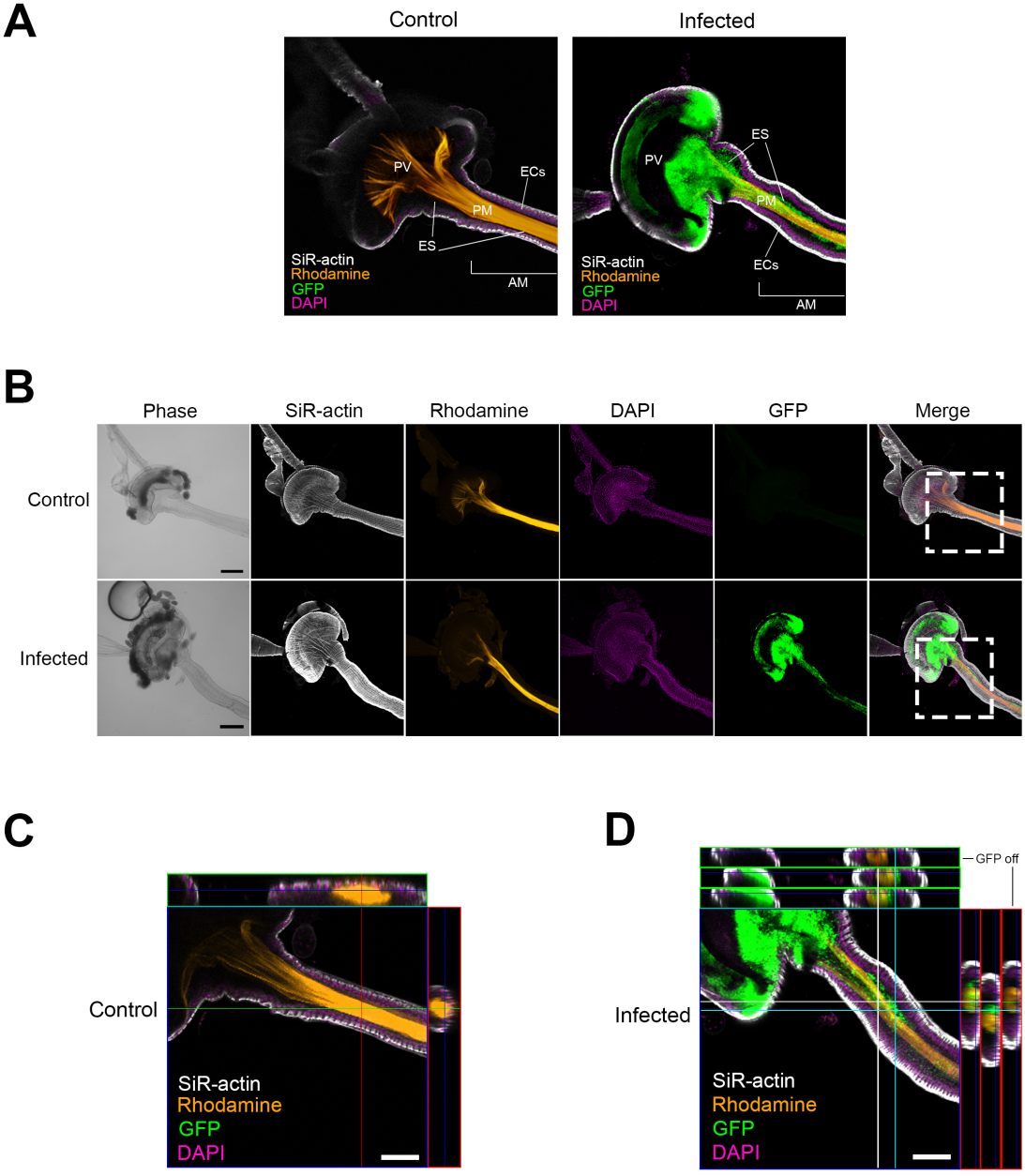


Figure 6.8. *T. brucei* localisation in PV and anterior midgut. (A) Examples of control and early-infected PV highlighting the areas of interest: PV, anterior midgut (AM), ectoperitrophic space (ES) and endothelial cells (EC). **(B)** Control and early-infected PV (5 d.p.i) with labelled actin (SiR-actin), PM (Rhodamine), nuclei (DAPI) and *T. brucei* (GFP). In the merge channel the dashed-line square highlights the area analyzed in C. **(C)** Orthogonal views of control and early-infected (5 d.p.i) PV. Infected PV has two orthogonal sections, and the outer one without GFP fluorescence. Scale bars = 100 μm.

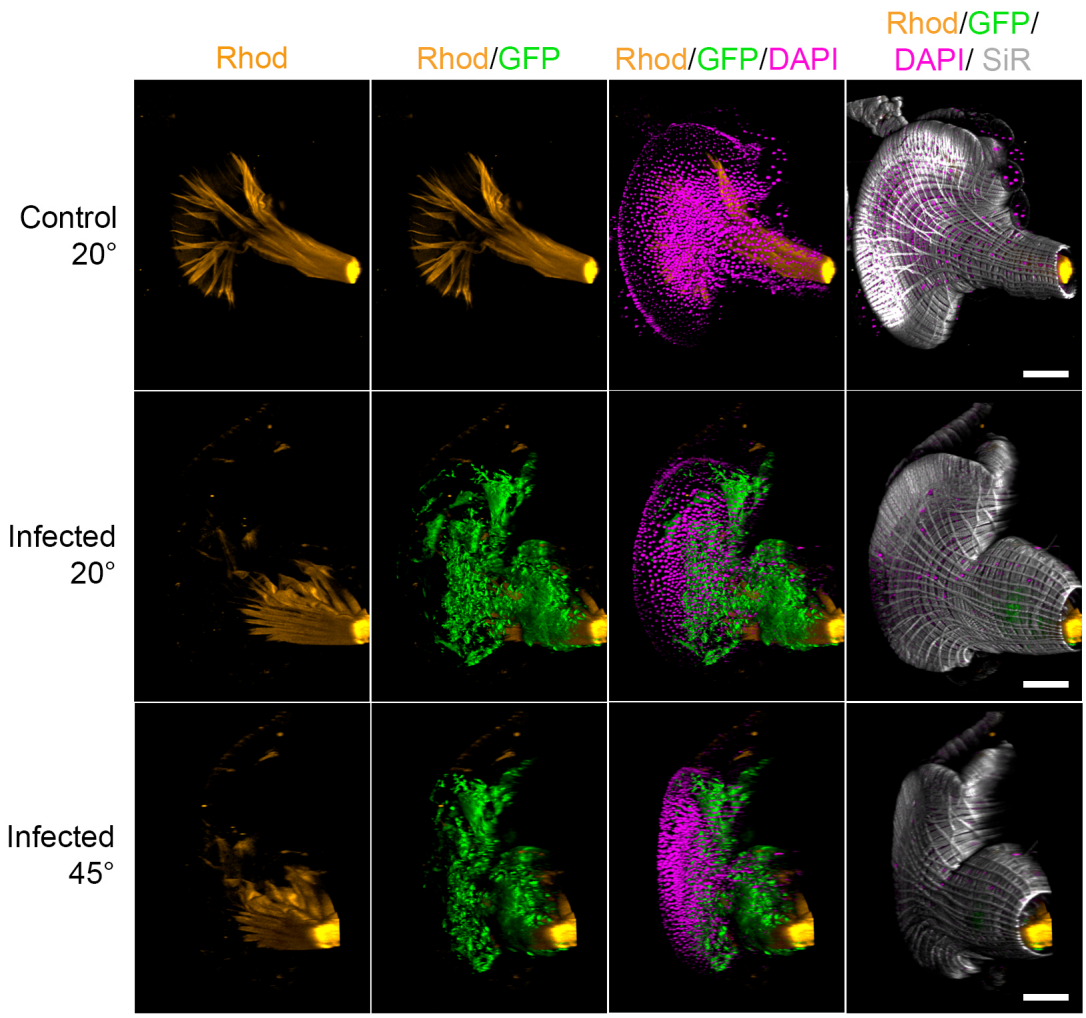


Figure 6.9. 3D reconstruction of infected proventriculus. Sequentially added fluorescent channels of control and early-infected PV. PM stained with WGA-rhodamine (Rhod), *T. brucei* (GFP), nuclei (DAPI) and actin (SiR). Reconstructions rotated at 20° and 45° clockwise to facilitate visualisation. Scale bars: 100 μm.

6.3.2.3 Imaging the tsetse *T. brucei*-infected PM *ex vivo*

Since in the orthogonal views of infected PV and AM it was observed that parasites formed small denser areas in the ES, the PM was dissected out and analysed on its own *ex vivo*. It could be observed that many parasites remained trapped in the PM after washes, suggesting they could be attached or trapped between PM layers. After PM staining with WGA and DAPI staining of nuclei along

with GFP fluorescence from parasites (**Figure 6.10-A**), it could be observed that the high-density areas corresponded to multiple parasites compartmentalised within the PM layer in a 'pocket'-like structure. Several pockets were 3D reconstructed and their corresponding orthogonal views showed that indeed the parasites are trapped between PM layers (**Figure 6.10-B**). Considering the thickness of the fluorescence surrounding the pocket, it could be suggested that the parasites are trapped between layers 1 (thinnest layer facing the Pm lumen) and 2-plus-3 (thickest layer facing the ES). This was further confirmed in TEM images of MG sections (in preparation). In **Figure 6.10-C**, a schematic of an infected MG and PM sections simplifies the visualisation of parasite pockets in the PM.

Besides finding parasites in pockets, single ones could also be found across the isolated PM. Using 3D reconstructions from z-stacks on single parasites it could be observed that some are attached either on the luminal or the ES side of the PM, and others are trapped between PM layers too (**Figure 6.11-A**). In a few reconstructions some parasites were seen in a 'PM crossing-like' event as the cells were transversally inserted in the PM (**Figure 6.11-A**) or between PM layers but not completely (**Figure 6.11-B**). It is difficult to prove they are crossing the three PM layers or only layers 2 and 3 which seem to be softer. It seems clear that the parasites found in the *ex vivo* stained PM are actively attached to it by some kind of interactive mechanism since they remain with the PM after many harsh washing steps. In addition, in most of the cases, the PM adopts the shape of the trypanosome attached to it. The biological role of this attachment to the PM remains unclear. Parasites found within PM layers either completely or partially trapped, or in a crossing-like event, could be either the result of being trapped in the PV where PM layers bind together, or that parasites can actively insert themselves mechanically within PM layers.

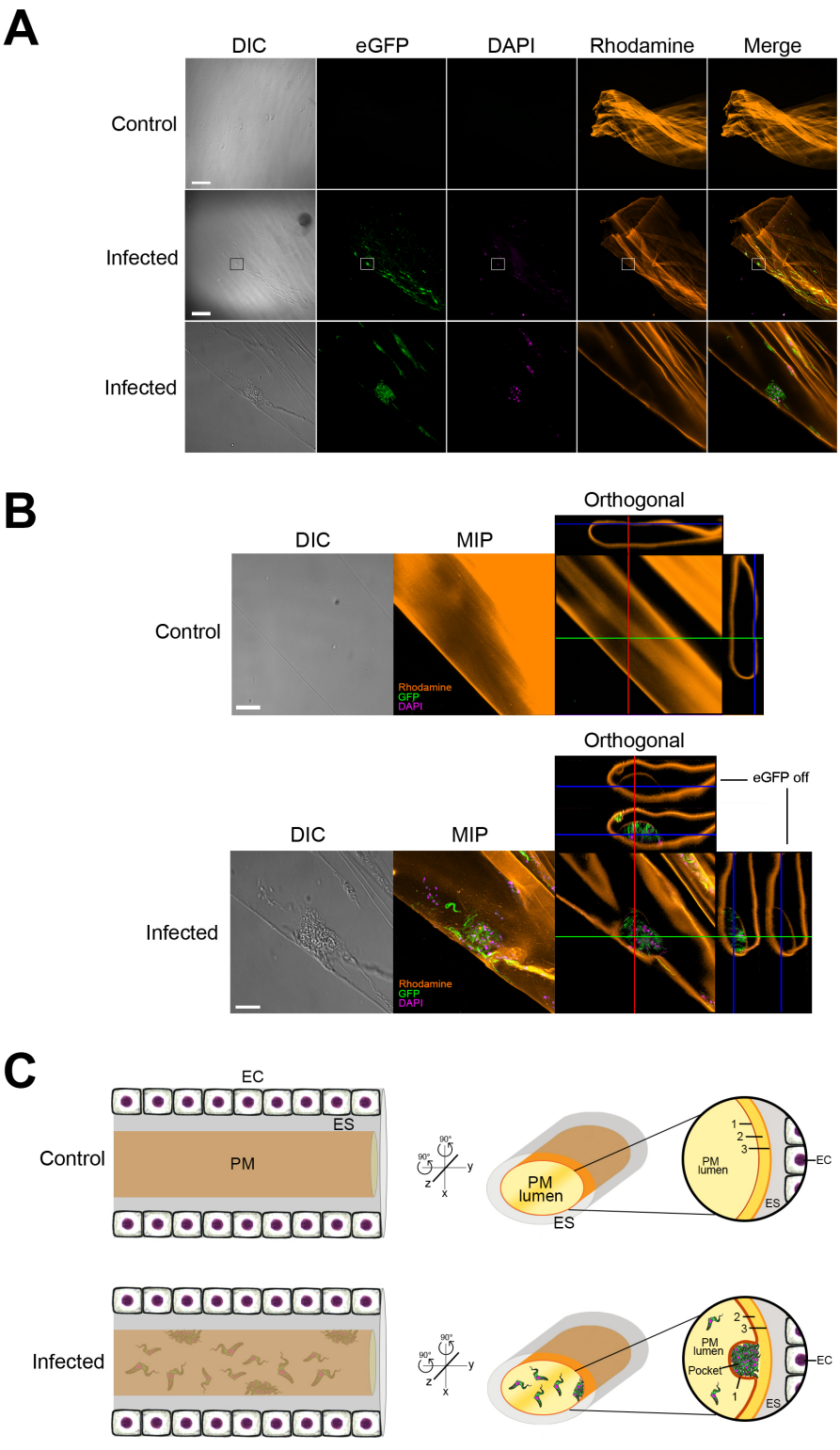


Figure 6.10. *T. brucei* trapped in PM 'pockets'. (A) Control and *T. brucei*-infected (GFP) PM sections stained with WGA-rhodamine and DAPI, and high magnification image of a pocket; scale bars = 200 μm. **(B)** Orthogonal reconstruction of control and infected (pocket) PM sections stained with WGA-rhodamine and DAPI, *T. brucei*

express GFP; differential interference contrast (DIC), maximum intensity projection of z-stacks (MIP); scale bars = 20 μm. **(C)** Schematic of top views of control and infected midgut sections (left, grey cylinder) showing the endothelial cells (white cells with purple nuclei), PM (orange cylinder) and ES (grey area). Sections rotate 90 degrees clockwise on the X and Y axes (middle) showing the PM from the side. Detail of the PM edge containing a *T. brucei* pocket between layers 1 and 2-plus-3 (right).

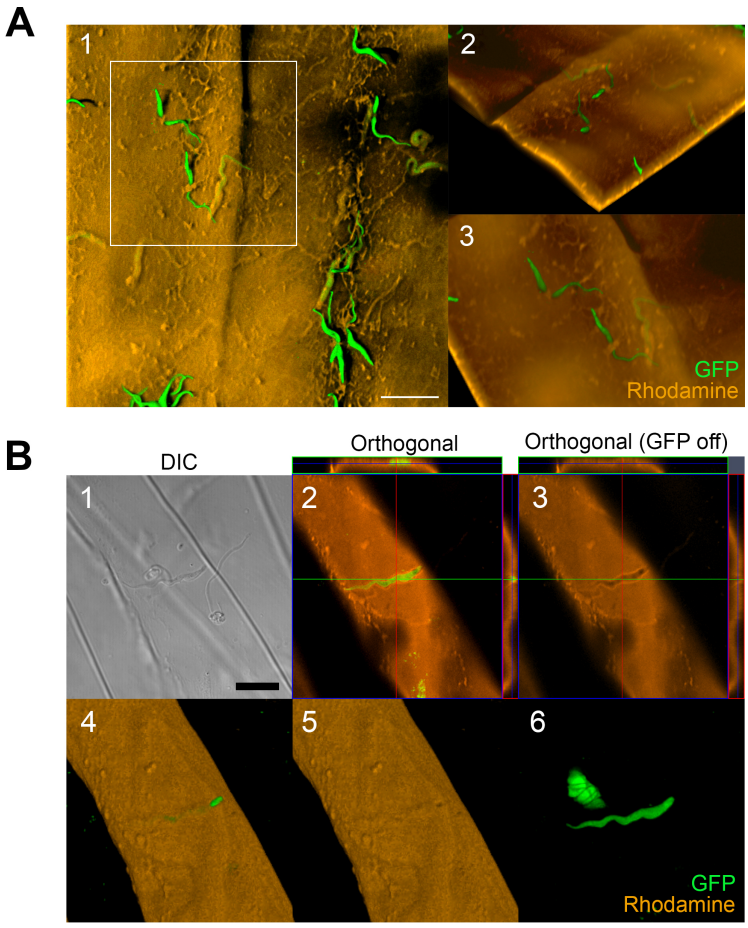


Figure 6.11. Single *T. brucei* cells interacting with the tsetse PM. (A) 3D reconstruction of an infected PM *ex vivo* with attached parasites in both luminal (L) and ES sides. An area with two parasites in a 'crossing-like' event is highlighted with a white rectangle (**1**). Two different perspectives of the two crossing parasites (**2, 3**) seen from the ES side. **(B)** A single parasite trapped within PM layers by DIC (**1**), maximum intensity projection and orthogonal view of GFP and rhodamine channels (**2**), rhodamine only (**3**), 3D reconstruction of the z-stacks for both channels (**4**), rhodamine only (**5**), GFP only (**6**). Scale bar = 5 μm

6.4. Discussion

6.4.1 New imaging methodology to reconstruct tsetse tissues

Good quality imaging methods are essential in basic research to provide direct evidence of biological phenomena. Although a broad range of imaging methods exists, they sometimes require optimisation to be applied to specific models of study. Here, I have set up a methodology based on laser scanning confocal microscopy to help visualise structural changes in tsetse tissues or during the interactions between *T. brucei* and tsetse fly tissues. The same methodology could also be applied to other tsetse and African trypanosome species such as *T. congolense*, and even to other vector-pathogen systems.

The structural staining relies on a commercial small drug conjugated to a far-red fluorophore, with high affinity for actin which is very conserved across cell types and organisms. This makes the structural staining method suitable for any system, although the resultant staining pattern is unpredictable. Thanks to the confocal technology, the fluorescence from the specimen can be captured only at the focal plane, allowing the acquisition of series of z-stack images across the thickness of the specimen and further faithful 3D reconstruction with fine detail. The 3D reconstructions can ease visualisation of biological interactions by observing from different angles, and the orthogonal views resulting from a combination of the z-stacks allow the correct identification of an object's position within the Z axis which is important, for example, to determine whether an object is either inside or outside a bigger one. The structural actin staining can be combined with more fluorescent markers to label different structures from both vector and pathogen. In the examples shown in the chapter, it was of great importance to visualise the nuclei of tsetse and trypanosome cells to identify DNA damage due to nitisinone exposure or for the identification of parasite stages, respectively. DAPI was chosen to stain DNA since it is easy to prepare, does not require cell permeabilisation and provides a strong clear signal in the blue range of light. There are other options available to consider, especially if live imaging for long periods of time is required, as DAPI intercalates into DNA strands and eventually leads to cell death. For instance, the commercial dyes DRAQ-5/7 allow DNA staining with high survival rates, although they emit light at similar wavelengths than SiR-actin which may give spectra overlapping issues. Additionally, the staining of the tsetse PM was of great importance to understand the

trypanosome infection dynamics in the MG and PV. The use of the fluorescent-tagged WGA lectin allowed the specific labelling of the PM (Schuster et al., 2017) since this is rich in chitin which is formed by long chains of N-acetylglucosamine residues (Lehane et al., 1996; Rose et al., 2014). Because the staining was intended for live imaging and the PM is constantly secreted by the PV (Lehane and Msangi, 1991), it was required to feed the tsetse with the lectin and dissect it out within a short period of time. PM staining within the PV is otherwise lost from 30 minutes after feeding the lectin. Lastly, a good fluorescence detection of trypanosomes was essential for the PV infectivity studies shown in this chapter. The expression of fluorescent proteins (FPs) by the parasite was the most convenient option for these experiments to avoid long immunostaining procedures and facilitate live imaging. The most used *T. brucei* cell line was the strain J-10 previously used by the community (Gibson and Bailey, 2003), which strongly expresses eGFP under a procyclin promoter. Alternatively, I also created a *T. brucei* AnTat 1.1 90:13 cell line that expresses mNeonGreen under tetracycline regulation and driven by the same promoter. The mNeonGreen FP has improved brightness (~3-fold brighter than eGFP) and quantum yield (33% higher than eGFP) (Shaner et al., 2013), which has helped to overcome poor imaging of parasites inside more opaque tissue tissues. The imaging of FPs, however, becomes challenging when fixation of the specimen is required. Most standard fixation methods rapidly quench FP fluorescence as a result of small structural changes in the chromophore. It was required to optimise fixation of FP-expressing trypanosomes to allow imaging. Soft fixation with PFA under cold conditions was found to be optimal to preserve FP fluorescence. Alternatively, immunostaining with antibodies against surface proteins such as procyclins could be used to fluorescently label the parasites, although this is unsuitable for trypanosomes trapped within the PM as antibodies cannot penetrate through its layers.

6.4.2. Assessing tsetse tissue damage upon nitisinone exposure

As previously described, the drug nitisinone, which is used to treat type-I tyrosinemia in humans, was shown to be very effective at killing *Rhodnius prolixus* (Sterkel et al., 2016), the vector of Chagas disease (Coura, 2015). It is suggested that nitisinone only affects hematophagous insects, so it can be used in the field to block transmission of virtually any vector-borne disease. Our group has recently shown that nitisinone is also effective in killing *G. m. morsitans*, the vector of sleeping sickness

(in preparation). In an attempt to further understand the cellular and molecular mechanisms that lead to the death of the tsetse upon exposure to nitisinone, I applied the developed method of 3D fluorescence tissue reconstruction to flies treated with the drug to search for specific patterns of tissue and cellular damage. In order to compare nitisinone-induced damaged at either “early” or the “late” stage, but also to perform tissue isolation and processing at the same time to prevent batch bias, flies were provided with the same dose of nitisinone mixed with different solutions (serum or medium), which was previously shown to lead to sub-lethal and lethal phenotypes, respectively. Overall, the “late” phenotype flies presented advanced damage of tissues that lost structural integrity and functionality. Actin and DAPI stainings were overall poor suggesting that cells lost actin and DNA content or condensation, probably due to cell death events. Imaging conditions required to be pushed to the limit to acquire low quality results. On the other hand, although actin and DAPI staining were somehow compromised in the early phenotype, it provided better results since all the analysed tissues conserved more or less their integrity, which allowed the examination of tissue and cellular damage in greater detail. For example, the salivary glands and Malpighian tubules are among the less affected tissues that could be analysed. Apparently unaffected MT was a surprising result since nitisinone quickly blocks excretion, suggesting that tissue integrity may not be the cause of excretion failure. The fat bodies were also slightly affected and only minor variance in shape and size could be appreciated. Remarkably, a few still conserved their structure in late phenotype flies although looked severely compromised. The digestive tract (MG and PV) was shown to be severely damaged in “early” phenotype flies and completely disintegrated in the “late” ones. Not only the MG structure itself was affected but also the transversal muscle fibres seemed to have collapsed. Lastly, flight muscle fibres were also found to be highly affected upon nitisinone treatment, especially in “late” phenotype flies. Despite not revealing a precise mechanism by which the tsetse dies upon nitisinone ingestion, these series of imaging analyses helped to understand the scope of tissue damage in detail, which could not be determined by using conventional microscopy techniques. Further analyses to elucidate the molecular mechanisms of death induced by nitisinone are on their way, including untargeted metabolomics.

6.4.3. Studies on a novel alternative route of tsetse PV infection by *Trypanosoma brucei*

The methods presented in this chapter were of great usefulness to investigate a novel tropism of *T. brucei* in the tsetse. This observation describes an alternative route of early gut colonisation through the PV, which challenges the current understanding on how the parasite establishes an infection in the tsetse vector. It is thought that *T. brucei* first invades the ES from the MG by crossing the PM, and then it migrates to the PV and establishes an infection there. The alternative route suggests the parasites invade the PV as early as 3 days post infection, and parasites can access the ES from there or get trapped within PM layers since the PM is continuously synthesised in the PV. For both proving the hypothesis and describing the mechanism it was required to develop tailored imaging methods to visualise and determine the exact location of the parasites inside tsetse tissues with little alterations. After staining and mounting optimisations, the method allowed to image trypanosomes both within the MG and PV, and with the PM and ES well defined inside. This methodology complements recent imaging techniques of trypanosomes within the tsetse midgut (Schuster et al., 2017), although with simpler and quicker methodologies, and suitable to be used in any confocal microscope.

The early infection of the PV was first found by live confocal microscopy and detailed analyses showed that PVs contain normal proventricular parasite forms, except for the absence of epimastigote forms which appear to develop later. Experiments with tsetse showed that the tropism is shared by at least 3 different *T. brucei* strains, but it is lost when parasites are kept in culture for a long period of time. This also appears to have an impact in the transmission index, suggesting that an early PV infection may modulate parasite transmissibility.

When analysing the localisation of trypanosomes in the reconstructed PV and AM with PM staining, it could be observed that they invade the ES in both PV and AM, and the distribution in those spaces was not always homogeneous but in high-density patches. In addition, when the infected PM was analysed *ex vivo*, it could be observed that contained multiple trypanosomes attached to it on both sides of the matrix, but also trapped in between PM layers in cyst-like structures or ‘pockets’. Further analyses of multiple TEM images of infected PM revealed that parasites are always trapped between PM layers 1 and 2, although no PM crossing events were

ever observed by us or any other group. However, in a few confocal reconstructions of infected PMs it could be observed parasites partially trapped in the PM in crossing-like events. It remains unknown whether these parasites actively crossed the three PM layers, or are only able to bury themselves within layers 2 and 3 (opposite to the midgut lumen). Alternatively, these parasites could have been trapped in the PV at the PM assembling site and dragged away because of PM excretion.

Altogether, these imaging methods combined with TEM allowed us to define a novel tropism for the PV at early stages of infection, which is an alternative way for *T. brucei* to invade the ectoperitrophic space of the tsetse gut, and it is likely to modulate parasite transmission.

CHAPTER 7

Overall discussion, reflections and future perspectives

This thesis first describes how a proteomics approach served for the identification of a new family of hypothetical *T. brucei* proteins found in salivary gland-dwelling parasites. Based on *in silico* searches, the five predicted MISP isoforms in *T. brucei* can be GPI-anchored surface proteins, although this remain to be biochemically proven. Subsequent experimental work showed that *misp* expression is developmentally regulated, but mainly transcribed in salivary gland stages. This is similar to the expression pattern observed in BARP proteins in the fly. Thanks to the creation of a polyclonal antibody, MISP could be seen mostly expressed by metacyclic cells and localised evenly throughout the cell surface, in great contrast with the decreasing expression of BARP (Urwyler et al., 2007). Strikingly, live immunostaining of metacyclics showed that MISP epitopes were accessible to antibodies which suggests that may not be completely masked by the VSG coat. Further resolution of the MISP crystal structure and molecular modelling supported this hypothesis. While the core N-termini adopt a triple helical bundle structure, which highly resembles that of other unrelated trypanosome surface proteins, the random coiled C-termini with no fixed secondary structure of some isoforms is significantly longer than those in VSG and BARP, and therefore it may project the core over the VSG coat. More insights into the full structure of MISP could be obtained by performing NMR and small-angle X-ray scattering (SAXS) studies on the recombinant C-termini domains produced in this thesis. Altogether, these studies suggest that the glycocalyx of MCF cells is more permissive to the presence and exposure of invariant GPI-anchored proteins than in BSFs, probably due to the fact that MCF are not under pressure of the humoral immune system during the short time they reside in the mammalian host. The biological function of MISP cannot be inferred from *in silico* searches and it currently remains unknown. It can be hypothesised that they may be surface receptors since they display a hydrophobic molecular pocket at the distal end which may coordinate a ligand molecule. Interestingly, the residues around the pocket present the highest divergence among isoforms suggesting they may coordinate different ligands, potentially in different hosts. The results of the pulldown assays presented in this thesis may identify a putative partner in tsetse saliva, although more MISP isoforms should be screened in the future to get a more comprehensive outcome. This putative ligand may not necessarily be soluble but a surface molecule on a host cell for cell recognition or adhesion, including the possibility of MISP being actively shed as the high density of protease recognition sites at the C-termini suggests.

In this thesis I also attempted to determine the essentiality of MISP for parasite development in the tsetse using RNAi. Although gene knockdown was demonstrated in SG parasites for the first time, the absence of a clear phenotype and the mild gene down-regulation do not confidently prove that MISP is not important for parasite development in SGs. Gene knockout approaches would provide clear answers, although classic knockout by cassette replacement is virtually unfeasible due to the presence of unrelated genes in between *misp*, as previously explained. The recent CRISPR-Cas9 technology should be used to provide a technical solution to this problem, as it would specifically target only *misp* genes and without the need of using multiple selection markers. Tsetse infections with these knocked-out parasites would provide more insights into function of MISP in either development or during transmission. Furthermore, MISP are the ideal surface marker to determine the kinetics of transformation of MCFs into BSFs after transmission by an infected tsetse. Additional mice work would include immunisations with recombinant MISP to determine the potential protective role of MISP as a vaccine. Lastly, the creation of *T. brucei* BSF cells ectopically expressing MISP could become a powerful tool for the study of some aspects of MISP, including the measurement of the relative abundance of MISP in relation to VSGs, the localisation of specific MISP isoforms or determining the MISP *N*-glycan structures.

In order to continue our studies on MISP, a different approach was taken for increasing the yield of MCFs. A previously reported *in vitro* system for the generation of MCF upon overexpression of RBP6 (Kolev et al., 2012) was successfully set up and new aspects of it were characterised. Although differentiation was triggered significantly earlier than previously reported (likely due to differences in the expression levels of RPB6), the overall efficiency in the production of MCF and other stages was lower. Remarkably, I showed that differentiation events have an impact in cell growth speed and trigger the expression of MISP and mVSGs. Preliminary experiments also suggest that RBP6 promotes differentiation of MCFs into BSF-like cells. This was suggested by the presence of cells with a similar shape to MCFs, but bigger and having 2 nuclei and 2 kDNAs, and expressing CRD epitopes on the surface. Further investigations need to validate this novel stage formed during *in vitro* differentiation by detecting, for instance, expression of transferrin receptor (only expressed in BSFs) and by looking at the activation of canonical BES (Cross et al., 2014). No further studies were carried out on these *in vitro* metacyclics due to the low yield of production, although this could be overcome in the future by using the ion

exchange purification step from large scale cultures. Additionally, I did some functional studies on RBP6 to understand its biological role during development in the tsetse, showing that it is mainly expressed in PV and SG parasites and that it appears not to be essential for development in the tsetse, although this should be further corroborated using conditional-null mutants.

In pursuit of the identification of genes important for promoting salivary gland infections, a comparative transcriptomics analysis was performed on two *T. brucei* strains. These strains differ in the ability to infect tsetse flies as one (Lister 427, 29:13 clone) is unable to establish salivary gland infections. This revealed several genes involved in folate transport, amino acid metabolism and RNA binding, which were shown to be up-regulated in tsetse late infection stages and were further validated by overexpression in the deficient strain. None of the candidates altered the infectivity phenotype except for a hypothetical cytosolic protein with a predicted MYND-like zinc finger domain. This protein was found to be up-regulated in late proventricular infections in the competent strain only, and its overexpression led to the recovery of SG infections. The function of this hypothetical protein remains unknown and cannot be inferred only from the presence of the zinc finger domain, which has been found to be involved in very different binding functionalities. More studies should be carried out to understand the role of this hypothetical protein in parasite development, including gene knockout in a fly-transmissible strain and immunoprecipitation to identify interactive RNA and protein partners.

The cell surface of *T. brucei* is known to be heavily *N*-glycosylated, including the well-studied VSGs and EP procyclins (Ferguson, 1997). Other major surface proteins such as BARP or MISP, are also predicted to be *N*-glycosylated, especially considering that *T. brucei* OSTs are very processive enzymes. During this thesis, I attempted to collect MCF cells from multiple tsetse infections with the aim of performing mass spectrometry-based glycoproteomics for the structural determination of MCF *N*-glycans. Despite having set up a standard methodology for the tsetse infection and collection of MCFs, I failed to obtain enough number of cells to perform glycan structural analyses. However, from the functional point of view, I found that the OST STT3B isoform, which is the major contributor of high mannose *N*-glycans in BSFs and PCFs, is the only isoform detected in parasites from the tsetse midgut, proventriculus and salivary glands. However, it cannot be ruled out that low expression of other isoforms was missed in the analyses and their activity may be

reflected in changes in *N*-glycan structures in certain stages. More sensitive analyses could be performed in the future using higher amounts of tsetse-derived parasites or even using RBP6 cells, although the latter is the less preferred option as the correct developmental expression of STT3s may not follow the fast differentiation occurring *in vitro*. STT3 expression measurements by real-time RT-PCR are virtually unfeasible due to the high sequence conservation between isoforms. Studying the essentiality of *N*-glycosylation for the parasite during development in the tsetse was of great importance too. Despite failing in the creation of conditional-null mutants defective in parasite *N*-glycosylation, RNAi of *stt3* genes showed a decrease in tsetse infectivity, similarly to previous reports (Izquierdo et al., 2009b). Infectivity was lower in the proventriculus compared with the midgut, and parasites were completely unable to infect the tsetse SG. These results point at any essential enzyme in the *N*-glycosylation pathway as a target to explore in drug discovery approaches.

Lastly, I set up a method to reconstruct tsetse tissues in 3D for multiple purposes. Based on a combination of different staining methods using fluorescent probes, the methods proved to be useful for the study of insect tissue damage and the interactions between trypanosomes and tsetse. First, it was observed that the tyrosine metabolism inhibitor drug, nitisinone, causes tissue damage as shown by the total collapse of muscle fibres from different organs and tissues. Further investigations could include using other fluorescent molecular probes for determining, for instance, if the tissue destruction triggered by nitisinone occurs via apoptosis, necrosis or oxidative stress. Second, a novel tropism of *T. brucei* in the tsetse could be studied in more detail. *T. brucei* can invade the tsetse proventriculus earlier than expected, possibly as a parallel route for reaching the midgut ectoperitrophic space. Future work to elucidate the role of this tropism in transmission will also benefit from these methods, which could also be applied to study late stages of infection including those in the salivary glands or widely to study vector-pathogen interactions in other models.

In summary, the work carried out in this thesis has opened the door to study MISP as a vaccine candidate against sleeping sickness and potentially AAT caused by *T. b. brucei, T. evansi* and *T. congolense*, in addition to be a good candidate for xenodiagnoses in the field. It has also identified a new hypothetical protein with a MYND zinc finger domain, which may have importance in development of *T. brucei* in the tsetse. In addition, it has reported the developmental expression pattern of the oligosaccharyltransferase isoforms in the *T. brucei* and the potential essentiality of

parasite *N*-glycosylation in the tsetse stages. Lastly, it has also contributed for the development of novel imaging methodologies to study vector-parasite interactions, which are currently lacking in the field.

REFERENCES

Abel, P.M., Kiala, G., Loa, V., Behrend, M., Musolf, J., Fleischmann, H., Theophile, J., Krishna, S., and Stich, A. (2004). Retaking sleeping sickness control in Angola. Trop Med Int Health *9*, 141-148.

Acosta-Serrano, A., Cole, R.N., and Englund, P.T. (2000). Killing of *Trypanosoma brucei* by concanavalin A: structural basis of resistance in glycosylation mutants. J Mol Biol *304*, 633-644.

Acosta-Serrano, A., Cole, R.N., Mehlert, A., Lee, M.G., Ferguson, M.A., and Englund, P.T. (1999). The procyclin repertoire of *Trypanosoma brucei*. Identification and structural characterization of the Glu-Pro-rich polypeptides. J Biol Chem *274*, 29763-29771.

Acosta-Serrano, A., Vassella, E., Liniger, M., Kunz Renggli, C., Brun, R., Roditi, I., and Englund, P.T. (2001). The surface coat of procyclic *Trypanosoma brucei*: programmed expression and proteolytic cleavage of procyclin in the tsetse fly. Proc Natl Acad Sci U S A *98*, 1513-1518.

Aksoy, E., Vigneron, A., Bing, X., Zhao, X., O'Neill, M., Wu, Y.N., Bangs, J.D., Weiss, B.L., and Aksoy, S. (2016). Mammalian African trypanosome VSG coat enhances tsetse's vector competence. Proc Natl Acad Sci U S A *113*, 6961-6966.

Alarcon, C.M., Son, H.J., Hall, T., and Donelson, J.E. (1994). A monocistronic transcript for a trypanosome variant surface glycoprotein. Mol Cell Biol *14*, 5579-5591.

Alsford, S., Kelly, J.M., Baker, N., and Horn, D. (2013). Genetic dissection of drug resistance in trypanosomes. Parasitology *140*, 1478-1491.

Alves-Silva, J., Ribeiro, J.M., Van Den Abbeele, J., Attardo, G., Hao, Z., Haines, L.R., Soares, M.B., Berriman, M., Aksoy, S., and Lehane, M.J. (2010). An insight into the sialome of *Glossina morsitans morsitans*. BMC Genomics *11*, 213.

Apweiler, R., Attwood, T.K., Bairoch, A., Bateman, A., Birney, E., Biswas, M., Bucher, P., Cerutti, L., Corpet, F., Croning, M.D.*, et al.* (2001). The InterPro database, an integrated documentation resource for protein families, domains and functional sites. Nucleic Acids Res *29*, 37-40.

Archer, S.K., Luu, V.D., de Queiroz, R.A., Brems, S., and Clayton, C. (2009). *Trypanosoma brucei* PUF9 regulates mRNAs for proteins involved in replicative processes over the cell cycle. PLoS Pathog *5*, e1000565.

Arrese, E.L., and Soulages, J.L. (2010). Insect fat body: energy, metabolism, and regulation. Annu Rev Entomol *55*, 207-225.

Ashkenazy, H., Abadi, S., Martz, E., Chay, O., Mayrose, I., Pupko, T., and Ben-Tal, N. (2016). ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. Nucleic Acids Res *44*, W344-350.

Ashkenazy, H., Erez, E., Martz, E., Pupko, T., and Ben-Tal, N. (2010). ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. Nucleic Acids Res *38*, W529-533.

Atayde, V.D., Aslan, H., Townsend, S., Hassani, K., Kamhawi, S., and Olivier, M. (2015). Exosome Secretion by the Parasitic Protozoan *Leishmania* within the Sand Fly Midgut. Cell Rep *13*, 957-967.

Atrih, A., Richardson, J.M., Prescott, A.R., and Ferguson, M.A. (2005). *Trypanosoma brucei* glycoproteins contain novel giant poly-N-acetyllactosamine carbohydrate chains. J Biol Chem *280*, 865-871.

Auty, H., Anderson, N.E., Picozzi, K., Lembo, T., Mubanga, J., Hoare, R., Fyumagwa, R.D., Mable, B., Hamill, L., Cleaveland, S.*, et al.* (2012). Trypanosome diversity in wildlife species from the serengeti and Luangwa Valley ecosystems. PLoS Negl Trop Dis *6*, e1828.

Barrett, M.P. (2006). The rise and fall of sleeping sickness. Lancet *367*, 1377-1378.

Barrett, M.P., Burchmore, R.J., Stich, A., Lazzari, J.O., Frasch, A.C., Cazzulo, J.J., and Krishna, S. (2003). The trypanosomiases. Lancet *362*, 1469-1480.

Barry, J.D., Crowe, J.S., and Vickerman, K. (1983). Instability of the *Trypanosoma brucei rhodesiense* metacyclic variable antigen repertoire. Nature *306*, 699-701.

Bartossek, T., Jones, N.G., Schafer, C., Cvitkovic, M., Glogger, M., Mott, H.R., Kuper, J., Brennich, M., Carrington, M., Smith, A.S.*, et al.* (2017). Structural basis for the shielding function of the dynamic trypanosome variant surface glycoprotein coat. Nat Microbiol *2*, 1523-1532.

Bayne, R.A., Kilbride, E.A., Lainson, F.A., Tetley, L., and Barry, J.D. (1993). A major surface antigen of procyclic stage *Trypanosoma congolense*. Mol Biochem Parasitol *61*, 295-310.

Beecroft, R.P., Roditi, I., and Pearson, T.W. (1993). Identification and characterization of an acidic major surface glycoprotein from procyclic stage *Trypanosoma congolense*. Mol Biochem Parasitol *61*, 285-294.

Belda, E., Moya, A., Bentley, S., and Silva, F.J. (2010). Mobile genetic element proliferation and gene inactivation impact over the genome structure and metabolic capabilities of *Sodalis glossinidius*, the secondary endosymbiont of tsetse flies. BMC Genomics *11*, 449.

Beneke, T., Madden, R., Makin, L., Valli, J., Sunter, J., and Gluenz, E. (2017). A CRISPR Cas9 high-throughput genome editing toolkit for kinetoplastids. R Soc Open Sci *4*, 170095.

Berriman, M., Ghedin, E., Hertz-Fowler, C., Blandin, G., Renauld, H., Bartholomeu, D.C., Lennard, N.J., Caler, E., Hamlin, N.E., Haas, B.*, et al.* (2005). The genome of the African trypanosome *Trypanosoma brucei*. Science *309*, 416-422.

Blum, J., Schmid, C., and Burri, C. (2006). Clinical aspects of 2541 patients with second stage human African trypanosomiasis. Acta Trop *97*, 55-64.

Blum, J.A., Zellweger, M.J., Burri, C., and Hatz, C. (2008). Cardiac involvement in African and American trypanosomiasis. Lancet Infect Dis *8*, 631-641.

Blum, M.L., Down, J.A., Gurnett, A.M., Carrington, M., Turner, M.J., and Wiley, D.C. (1993). A structural motif in the variant surface glycoproteins of *Trypanosoma brucei*. Nature *362*, 603-609.

Bottomley, C., Isham, V., Vivas-Martinez, S., Kuesel, A.C., Attah, S.K., Opoku, N.O., Lustigman, S., Walker, M., and Basanez, M.G. (2016). Modelling Neglected Tropical Diseases diagnostics: the sensitivity of skin snips for *Onchocerca volvulus* in near elimination and surveillance settings. Parasit Vectors *9*, 343.

Brenndorfer, M., and Boshart, M. (2010). Selection of reference genes for mRNA quantification in *Trypanosoma brucei*. Mol Biochem Parasitol *172*, 52-55.

Bronner, U., Doua, F., Ericsson, O., Gustafsson, L.L., Miezan, T.W., Rais, M., and Rombo, L. (1991). Pentamidine concentrations in plasma, whole blood and cerebrospinal fluid during treatment of *Trypanosoma gambiense* infection in Cote d'Ivoire. Trans R Soc Trop Med Hyg *85*, 608-611.

Buguet, A., Bourdon, L., Bouteille, B., Cespuglio, R., Vincendeau, P., Radomski, M.W., and Dumas, M. (2001). The duality of sleeping sickness: focusing on sleep. Sleep Med Rev *5*, 139-153.

Butikofer, P., Ruepp, S., Boschung, M., and Roditi, I. (1997). 'GPEET' procyclin is the major surface protein of procyclic culture forms of *Trypanosoma brucei brucei* strain 427. Biochem J *326 (Pt 2)*, 415-423.

Caljon, G., Van Reet, N., De Trez, C., Vermeersch, M., Perez-Morga, D., and Van Den Abbeele, J. (2016). The Dermis as a Delivery Site of *Trypanosoma brucei* for Tsetse Flies. PLoS Pathog *12*, e1005744.

Capewell, P., Cren-Travaille, C., Marchesi, F., Johnston, P., Clucas, C., Benson, R.A., Gorman, T.A., Calvo-Alvarez, E., Crouzols, A., Jouvion, G.*, et al.* (2016). The skin is a significant but overlooked anatomical reservoir for vector-borne African trypanosomes. Elife *5*.

Cardoso de Almeida, M.L., and Turner, M.J. (1983). The membrane form of variant surface glycoproteins of *Trypanosoma brucei*. Nature *302*, 349-352.

Casas-Sanchez, A., and Acosta-Serrano, A. (2016). Skin deep. Elife *5*.

Castillo-Acosta, V.M., Ruiz-Perez, L.M., Etxebarria, J., Reichardt, N.C., Navarro, M., Igarashi, Y., Liekens, S., Balzarini, J., and Gonzalez-Pacanowska, D. (2016). Carbohydrate-Binding Non-Peptidic Pradimicins for the Treatment of Acute Sleeping Sickness in Murine Models. PLoS Pathog *12*, e1005851.

Castillo-Acosta, V.M., Vidal, A.E., Ruiz-Perez, L.M., Van Damme, E.J., Igarashi, Y., Balzarini, J., and Gonzalez-Pacanowska, D. (2013). Carbohydrate-binding agents act as potent trypanocidals that elicit modifications in VSG glycosylation and reduced virulence in *Trypanosoma brucei*. Mol Microbiol *90*, 665-679.

Castro, O., Movsichoff, F., and Parodi, A.J. (2006). Preferential transfer of the complete glycan is determined by the oligosaccharyltransferase complex and not by the catalytic subunit. Proc Natl Acad Sci U S A *103*, 14756-14760.

Cecchi, G., Mattioli, R.C., Slingenbergh, J., and de la Rocque, S. (2008). Land cover and tsetse fly distributions in sub-Saharan Africa. Med Vet Entomol *22*, 364-373.

Chappuis, C.J., Beguin, S., Vlimant, M., and Guerin, P.M. (2013). Water vapour and heat combine to elicit biting and biting persistence in tsetse. Parasit Vectors *6*, 240.

Chappuis, F., Pittet, A., Bovier, P.A., Adams, K., Godineau, V., Hwang, S.Y., Magnus, E., and Buscher, P. (2002). Field evaluation of the CATT/*Trypanosoma brucei gambiense* on blood-impregnated filter papers for diagnosis of human African trypanosomiasis in southern Sudan. Trop Med Int Health *7*, 942-948.

Checchi, F., Filipe, J.A., Barrett, M.P., and Chandramohan, D. (2008a). The natural progression of Gambiense sleeping sickness: what is the evidence? PLoS Negl Trop Dis *2*, e303.

Checchi, F., Filipe, J.A., Haydon, D.T., Chandramohan, D., and Chappuis, F. (2008b). Estimates of the duration of the early and late stage of gambiense sleeping sickness. BMC Infect Dis *8*, 16.

Chen, X., Li, S., and Aksoy, S. (1999). Concordant evolution of a symbiont with its host insect species: molecular phylogeny of genus *Glossina* and its bacteriome-associated endosymbiont, *Wigglesworthia glossinidia*. J Mol Evol *48*, 49-58.

Cheng, Q., and Aksoy, S. (1999). Tissue tropism, transmission and expression of foreign genes in vivo in midgut symbionts of tsetse flies. Insect Mol Biol *8*, 125-132.

Cheng, Q., Ruel, T.D., Zhou, W., Moloo, S.K., Majiwa, P., O'Neill, S.L., and Aksoy, S. (2000). Tissue distribution and prevalence of *Wolbachia* infections in tsetse flies, *Glossina* spp. Med Vet Entomol *14*, 44-50.

Christiano, R., Kolev, N.G., Shi, H., Ullu, E., Walther, T.C., and Tschudi, C. (2017). The proteome and transcriptome of the infectious metacyclic form of *Trypanosoma brucei* define quiescent cells primed for mammalian invasion. Mol Microbiol.

Clayton, C. (2013). The regulation of trypanosome gene expression by RNA-binding proteins. PLoS Pathog *9*, e1003680.

Cordoba, M., Almeida, P., Goicolea, J., Andrade, I., Fraile, J., Rabago, P., and Rabago, G. (1983). Early in vivo hemodynamic evaluation of the mitral Medtronic-Hall cardiac valve. Thorac Cardiovasc Surg *31 Spec 2*, 85-88.

Coura, J.R. (2015). The main sceneries of Chagas disease transmission. The vectors, blood and oral transmissions--a comprehensive review. Mem Inst Oswaldo Cruz *110*, 277-282.

Courtin, F., Camara, M., Rayaisse, J.B., Kagbadouno, M., Dama, E., Camara, O., Traore, I.S., Rouamba, J., Peylhard, M., Somda, M.B.*, et al.* (2015). Reducing Human-Tsetse Contact Significantly Enhances the Efficacy of Sleeping Sickness Active Screening Campaigns: A Promising Result in the Context of Elimination. PLoS Negl Trop Dis *9*, e0003727.

Cova, M., Rodrigues, J.A., Smith, T.K., and Izquierdo, L. (2015). Sugar activation and glycosylation in *Plasmodium*. Malar J *14*, 427.

Cross, G.A. (1975). Identification, purification and properties of clone-specific glycoprotein antigens constituting the surface coat of *Trypanosoma brucei*. Parasitology *71*, 393-417.

Cross, G.A., Kim, H.S., and Wickstead, B. (2014). Capturing the variant surface glycoprotein repertoire (the VSGnome) of *Trypanosoma brucei* Lister 427. Mol Biochem Parasitol *195*, 59-73.

Cully, D.F., Ip, H.S., and Cross, G.A. (1985). Coordinate transcription of variant surface glycoprotein genes and an expression site associated gene family in *Trypanosoma brucei*. Cell *42*, 173-182.

De Kyvon, M.A., Maakaroun-Vermesse, Z., Lanotte, P., Priotto, G., Perez-Simarro, P., Guennoc, A.M., De Toffol, B., Paris, L., Bernard, L., Goudeau, A.*, et al.* (2016). Congenital Trypanosomiasis in Child Born in France to African Mother. Emerg Infect Dis *22*, 935-937.

De Lange, T., and Borst, P. (1982). Genomic environment of the expression-linked extra copies of genes for surface antigens of *Trypanosoma brucei* resembles the end of a chromosome. Nature *299*, 451-453.

Dean, S., Marchetti, R., Kirk, K., and Matthews, K.R. (2009). A surface transporter family conveys the trypanosome differentiation signal. Nature *459*, 213-217.

Dewar, S., Sienkiewicz, N., Ong, H.B., Wall, R.J., Horn, D., and Fairlamb, A.H. (2016). The Role of Folate Transport in Antifolate Drug Action in *Trypanosoma brucei*. J Biol Chem *291*, 24768-24778.

Duggan, A.J., and Hutchinson, M.P. (1966). Sleeping sickness in Europeans: a review of 109 cases. J Trop Med Hyg *69*, 124-131.

Emes, R.D., and Yang, Z. (2008). Duplicated paralogous genes subject to positive selection in the genome of *Trypanosoma brucei*. PLoS One *3*, e2295.

Engstler, M., Pfohl, T., Herminghaus, S., Boshart, M., Wiegertjes, G., Heddergott, N., and Overath, P. (2007). Hydrodynamic flow-mediated protein sorting on the cell surface of trypanosomes. Cell *131*, 505-515.

Engstler, M., Reuter, G., and Schauer, R. (1993). The developmentally regulated trans-sialidase from *Trypanosoma brucei* sialylates the procyclic acidic repetitive protein. Mol Biochem Parasitol *61*, 1-13.

Engstler, M., and Schauer, R. (1993). Sialidases from African trypanosomes. Parasitol Today *9*, 222-225.

Enyaru, J.C., Matovu, E., Nerima, B., Akol, M., and Sebikali, C. (2006). Detection of *T.b. rhodesiense* trypanosomes in humans and domestic animals in south east Uganda by amplification of serum resistance-associated gene. Ann N Y Acad Sci *1081*, 311-319.

Feasey, N., Wansbrough-Jones, M., Mabey, D.C., and Solomon, A.W. (2010). Neglected tropical diseases. Br Med Bull *93*, 179-200.

Fenech, M. (2012). Folate (vitamin B9) and vitamin B12 and their function in the maintenance of nuclear and mitochondrial genome integrity. Mutat Res *733*, 21-33.

Ferguson, M.A. (1997). The surface glycoconjugates of trypanosomatid parasites. Philos Trans R Soc Lond B Biol Sci *352*, 1295-1302.

Ferguson, M.A. (1999). The structure, biosynthesis and functions of glycosylphosphatidylinositol anchors, and the contributions of trypanosome research. J Cell Sci *112 (Pt 17)*, 2799-2809.

Ferguson, M.A., Homans, S.W., Dwek, R.A., and Rademacher, T.W. (1988). Glycosyl-phosphatidylinositol moiety that anchors *Trypanosoma brucei* variant surface glycoprotein to the membrane. Science *239*, 753-759.

Ferguson, M.A., and Williams, A.F. (1988). Cell-surface anchoring of proteins via glycosyl-phosphatidylinositol structures. Annu Rev Biochem *57*, 285-320.

Fevre, E.M., Wissmann, B.V., Welburn, S.C., and Lutumba, P. (2008). The burden of human African trypanosomiasis. PLoS Negl Trop Dis *2*, e333.

Fox, J.A., Duszenko, M., Ferguson, M.A., Low, M.G., and Cross, G.A. (1986). Purification and characterization of a novel glycan-phosphatidylinositol-specific phospholipase C from *Trypanosoma brucei*. J Biol Chem *261*, 15767-15771.

Franco, J.R. (2012). Monitoring the use of nifurtimox-eflornithine combination therapy (NECT) in the treatment of second stage gambiense human African trypanosomiasis. Research and Reports in Tropical Medicine *3*, 93-101.

Freymann, D., Down, J., Carrington, M., Roditi, I., Turner, M., and Wiley, D. (1990). 2.9 A resolution structure of the N-terminal domain of a variant surface glycoprotein from *Trypanosoma brucei*. J Mol Biol *216*, 141-160.

Gadelha, C., Zhang, W., Chamberlain, J.W., Chait, B.T., Wickstead, B., and Field, M.C. (2015). Architecture of a Host-Parasite Interface: Complex Targeting Mechanisms Revealed Through Proteomics. Mol Cell Proteomics *14*, 1911-1926.

Gibson, M.W., Dewar, S., Ong, H.B., Sienkiewicz, N., and Fairlamb, A.H. (2016). *Trypanosoma brucei* DHFR-TS Revisited: Characterisation of a Bifunctional and Highly Unstable Recombinant Dihydrofolate Reductase-Thymidylate Synthase. PLoS Negl Trop Dis *10*, e0004714.

Gibson, W., and Bailey, M. (2003). The development of *Trypanosoma brucei* within the tsetse fly midgut observed using green fluorescent trypanosomes. Kinetoplastid Biol Dis *2*, 1.

Gibson, W., Peacock, L., and Hutchinson, R. (2017). Microarchitecture of the tsetse fly proboscis. Parasit Vectors *10*, 430.

Gibson, W., and Stevens, J. (1999). Genetic exchange in the trypanosomatidae. Adv Parasitol *43*, 1-46.

Gibson, W.C. (2005). The SRA gene: the key to understanding the nature of *Trypanosoma brucei rhodesiense*. Parasitology *131*, 143-150.

Giordani, F., Morrison, L.J., Rowan, T.G., HP, D.E.K., and Barrett, M.P. (2016). The animal trypanosomiases and their chemotherapy: a review. Parasitology *143*, 1862-1889.

Giraud, E., Lestinova, T., Derrick, T., Martin, O., Dillon, R.J., Volf, P., Muller, I., Bates, P.A., and Rogers, M.E. (2018). *Leishmania* proteophosphoglycans regurgitated from infected sand flies accelerate dermal wound repair and exacerbate leishmaniasis via insulin-like growth factor 1-dependent signalling. PLoS Pathog *14*, e1006794.

Glisovic, T., Bachorik, J.L., Yong, J., and Dreyfuss, G. (2008). RNA-binding proteins and post-transcriptional gene regulation. FEBS Lett *582*, 1977-1986.

Gluenz, E., Wheeler, R.J., Hughes, L., and Vaughan, S. (2015). Scanning and three-dimensional electron microscopy methods for the study of *Trypanosoma brucei* and *Leishmania mexicana* flagella. Methods Cell Biol *127*, 509-542.

Glover, L., Hutchinson, S., Alsford, S., and Horn, D. (2016). VEX1 controls the allelic exclusion required for antigenic variation in trypanosomes. Proc Natl Acad Sci U S A *113*, 7225-7230.

Graf, F.E., Ludin, P., Wenzler, T., Kaiser, M., Brun, R., Pyana, P.P., Buscher, P., de Koning, H.P., Horn, D., and Maser, P. (2013). Aquaporin 2 mutations in *Trypanosoma brucei gambiense* field isolates correlate with decreased susceptibility to pentamidine and melarsoprol. PLoS Negl Trop Dis *7*, e2475.

Graham, S.V., and Barry, J.D. (1995). Transcriptional regulation of metacyclic variant surface glycoprotein gene expression during the life cycle of *Trypanosoma brucei*. Mol Cell Biol *15*, 5945-5956.

Grandgenett, P.M., Otsu, K., Wilson, H.R., Wilson, M.E., and Donelson, J.E. (2007). A function for a specific zinc metalloprotease of African trypanosomes. PLoS Pathog *3*, 1432-1445.

Grebaut, P., Melachio, T., Nyangmang, S., Eyenga, V.E., Njitchouang, G.R., Ofon, E., Njiokou, F., and Simo, G. (2016). Xenomonitoring of sleeping sickness transmission in Campo (Cameroon). Parasit Vectors *9*, 201.

Gross, C.T., and McGinnis, W. (1996). DEAF-1, a novel protein that binds an essential region in a Deformed response element. EMBO J *15*, 1961-1970.

Gruszynski, A.E., DeMaster, A., Hooper, N.M., and Bangs, J.D. (2003). Surface coat remodeling during differentiation of *Trypanosoma brucei*. J Biol Chem *278*, 24665-24672.

Guindon, S., and Gascuel, O. (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol *52*, 696-704.

Gunzl, A., Bruderer, T., Laufer, G., Schimanski, B., Tu, L.C., Chung, H.M., Lee, P.T., and Lee, M.G. (2003). RNA polymerase I transcribes procyclin genes and variant surface glycoprotein gene expression sites in *Trypanosoma brucei*. Eukaryot Cell *2*, 542-551.

Gustafsson, L. (1987). Handbook of drugs for tropical parasitic infections, 1st edn. Basingstoke: Taylor & Francis *160-63*.

Haines, L.R. (2002). PhD Thesis (University of Liverpool).

Haines, L.R. (2013). Examining the tsetse teneral phenomenon and permissiveness to trypanosome infection. Front Cell Infect Microbiol *3*, 84.

Haines, L.R., Lehane, S.M., Pearson, T.W., and Lehane, M.J. (2010). Tsetse EP protein protects the fly midgut from trypanosome establishment. PLoS Pathog *6*, e1000793.

Helenius, A., and Aebi, M. (2004). Roles of *N*-linked glycans in the endoplasmic reticulum. Annu Rev Biochem *73*, 1019-1049.

Helenius, J., Ng, D.T., Marolda, C.L., Walter, P., Valvano, M.A., and Aebi, M. (2002). Translocation of lipid-linked oligosaccharides across the ER membrane requires Rft1 protein. Nature *415*, 447-450.

Herwaldt, B.L. (2001). Laboratory-acquired parasitic infections from accidental exposures. Clin Microbiol Rev *14*, 659-688, table of contents.

Hese, K., Otto, C., Routier, F.H., and Lehle, L. (2009). The yeast oligosaccharyltransferase complex can be replaced by STT3 from *Leishmania major*. Glycobiology *19*, 160-171.

Higgins, M.K., Tkachenko, O., Brown, A., Reed, J., Raper, J., and Carrington, M. (2013). Structure of the trypanosome haptoglobin-hemoglobin receptor and implications for nutrient uptake and innate immunity. Proc Natl Acad Sci U S A *110*, 1905-1910.

Hira, P.R., and Husein, S.F. (1979). Some transfusion-induced parasitic infections in Zambia. J Hyg Epidemiol Microbiol Immunol *23*, 436-444.

Hirt, R.P., de Miguel, N., Nakjang, S., Dessi, D., Liu, Y.C., Diaz, N., Rappelli, P., Acosta-Serrano, A., Fiori, P.L., and Mottram, J.C. (2011). *Trichomonas vaginalis* pathobiology new insights from the genome sequence. Adv Parasitol *77*, 87-140.

Hirumi, H., and Hirumi, K. (1989). Continuous cultivation of *Trypanosoma brucei* blood stream forms in a medium containing a low concentration of serum protein without feeder cell layers. J Parasitol *75*, 985-989.

Holm, L., and Rosenstrom, P. (2010). Dali server: conservation mapping in 3D. Nucleic Acids Res *38*, W545-549.

Horn, D. (2014). Antigenic variation in African trypanosomes. Mol Biochem Parasitol *195*, 123-129.

Hwa, K.Y., Acosta-Serrano, A., Khoo, K.H., Pearson, T., and Englund, P.T. (1999). Protein glycosylation mutants of procyclic *Trypanosoma brucei*: defects in the asparagine-glycosylation pathway. Glycobiology *9*, 181-190.

Imhof, S., Vu, X.L., Butikofer, P., and Roditi, I. (2015). A Glycosylation Mutant of *Trypanosoma brucei* Links Social Motility Defects In Vitro to Impaired Colonization of Tsetse Flies In Vivo. Eukaryot Cell *14*, 588-592.

Ishemgulova, A., Hlavacova, J., Majerova, K., Butenko, A., Lukes, J., Votypka, J., Volf, P., and Yurchenko, V. (2018). CRISPR/Cas9 in *Leishmania mexicana*: A case study of LmxBTN1. PLoS One *13*, e0192723.

Izquierdo, L., Nakanishi, M., Mehlert, A., Machray, G., Barton, G.J., and Ferguson, M.A. (2009a). Identification of a glycosylphosphatidylinositol anchor-modifying beta1-3 *N*-acetylglucosaminyl transferase in *Trypanosoma brucei*. Mol Microbiol *71*, 478-491.

Izquierdo, L., Schulz, B.L., Rodrigues, J.A., Guther, M.L., Procter, J.B., Barton, G.J., Aebi, M., and Ferguson, M.A. (2009b). Distinct donor and acceptor specificities of *Trypanosoma brucei* oligosaccharyltransferases. EMBO J *28*, 2650-2661.

Jackson, A.P., Allison, H.C., Barry, J.D., Field, M.C., Hertz-Fowler, C., and Berriman, M. (2013). A cell-surface phylome for African trypanosomes. PLoS Negl Trop Dis *7*, e2121.

Jackson, C.H. (1946). An artificially isolated generation of tsetse flies (Diptera). Bull Entomol Res *37*, 291-299.

Jacobs, R.T., Nare, B., Wring, S.A., Orr, M.D., Chen, D., Sligar, J.M., Jenks, M.X., Noe, R.A., Bowling, T.S., Mercer, L.T.*, et al.* (2011). SCYX-7158, an orally-active benzoxaborole for the treatment of stage 2 human African trypanosomiasis. PLoS Negl Trop Dis *5*, e1151.

Jamonneau, V., Camara, O., Ilboudo, H., Peylhard, M., Koffi, M., Sakande, H., N'Dri, L., Sanou, D., Dama, E., Camara, M.*, et al.* (2015). Accuracy of individual rapid tests for serodiagnosis of gambiense sleeping sickness in West Africa. PLoS Negl Trop Dis *9*, e0003480.

Jamonneau, V., Garcia, A., Frezil, J.L., N'Guessan, P., N'Dri, L., Sanon, R., Laveissiere, C., and Truc, P. (2000). Clinical and biological evolution of human trypanosomiasis in C te d'Ivoire. Ann Trop Med Parasitol *94*, 831-835.

Jamonneau, V., Garcia, A., Ravel, S., Cuny, G., Oury, B., Solano, P., N'Guessan, P., N'Dri, L., Sanon, R., Frezil, J.L.*, et al.* (2002). Genetic characterization of *Trypanosoma brucei gambiense* and clinical evolution of human African trypanosomiasis in Cote d'Ivoire. Trop Med Int Health *7*, 610-621.

Jamonneau, V., Ilboudo, H., Kabore, J., Kaba, D., Koffi, M., Solano, P., Garcia, A., Courtin, D., Laveissiere, C., Lingue, K.*, et al.* (2012). Untreated human infections by *Trypanosoma brucei gambiense* are not 100% fatal. PLoS Negl Trop Dis *6*, e1691.

Jamonneau, V., Ravel, S., Koffi, M., Kaba, D., Zeze, D.G., Ndri, L., Sane, B., Coulibaly, B., Cuny, G., and Solano, P. (2004). Mixed infections of trypanosomes in tsetse and pigs and their epidemiological significance in a sleeping sickness focus of Cote d'Ivoire. Parasitology *129*, 693-702.

Jelk, J., Gao, N., Serricchio, M., Signorell, A., Schmidt, R.S., Bangs, J.D., Acosta-Serrano, A., Lehrman, M.A., Butikofer, P., and Menon, A.K. (2013). Glycoprotein biosynthesis in a eukaryote lacking the membrane protein Rft1. J Biol Chem *288*, 20616-20623.

Jinnelov, A., Ali, L., Tinti, M., Guther, M.L.S., and Ferguson, M.A.J. (2017). Single-subunit oligosaccharyltransferases of *Trypanosoma brucei* display different and predictable peptide acceptor specificities. J Biol Chem *292*, 20328-20341.

Johnson, P.J., Kooter, J.M., and Borst, P. (1987). Inactivation of transcription by UV irradiation of *T. brucei* provides evidence for a multicistronic transcription unit including a VSG gene. Cell *51*, 273-281.

Jordan, A. (1993). Tsetse-flies (Glossinidae). London: Chapman and Hall *333-88*.

Kagbadouno, M.S., Camara, M., Rouamba, J., Rayaisse, J.B., Traore, I.S., Camara, O., Onikoyamou, M.F., Courtin, F., Ravel, S., de Meeus, T.*, et al.* (2012). Epidemiology of sleeping sickness in Boffa (Guinea): where are the trypanosomes? PLoS Negl Trop Dis *6*, e1949.

Kariithi, H.M., Boeren, S., Murungi, E.K., Vlak, J.M., and Abd-Alla, A.M. (2016). A proteomics approach reveals molecular manipulators of distinct cellular processes in the salivary glands of *Glossina m. morsitans* in response to *Trypanosoma b. brucei* infections. Parasit Vectors *9*, 424.

Kelly, S., Reed, J., Kramer, S., Ellis, L., Webb, H., Sunter, J., Salje, J., Marinsek, N., Gull, K., Wickstead, B.*, et al.* (2007). Functional genomics in *Trypanosoma brucei*: a collection of vectors for the expression of tagged proteins from endogenous and ectopic gene loci. Mol Biochem Parasitol *154*, 103-109.

Kennedy, P.G. (2004). Human African trypanosomiasis of the CNS: current issues and challenges. J Clin Invest *113*, 496-504.

Kennedy, P.G. (2008). Diagnosing central nervous system trypanosomiasis: two stage or not to stage? Trans R Soc Trop Med Hyg *102*, 306-307.

Kolev, N.G., Ramey-Butler, K., Cross, G.A., Ullu, E., and Tschudi, C. (2012). Developmental progression to infectivity in *Trypanosoma brucei* triggered by an RNA-binding protein. Science *338*, 1352-1353.

Kornfeld, R., and Kornfeld, S. (1985). Assembly of asparagine-linked oligosaccharides. Annu Rev Biochem *54*, 631-664.

Kosinski, J., Mosalaganti, S., von Appen, A., Teimer, R., DiGuilio, A.L., Wan, W., Bui, K.H., Hagen, W.J., Briggs, J.A., Glavy, J.S*., et al.* (2016). Molecular architecture of the inner ring scaffold of the human nuclear pore complex. Science *352*, 363-365.

Krogh, A., Larsson, B., von Heijne, G., and Sonnhammer, E.L. (2001). Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J Mol Biol *305*, 567-580.

Laemmli, U.K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature *227*, 680-685.

Lane-Serff, H., MacGregor, P., Lowe, E.D., Carrington, M., and Higgins, M.K. (2014). Structural basis for ligand and innate immunity factor uptake by the trypanosome haptoglobin-haemoglobin receptor. Elife *3*, e05553.

Lane-Serff, H., MacGregor, P., Peacock, L., Macleod, O.J., Kay, C., Gibson, W., Higgins, M.K., and Carrington, M. (2016). Evolutionary diversification of the trypanosome haptoglobin-haemoglobin receptor from an ancestral haemoglobin receptor. Elife *5*.

Langley, R.W.a.P.A. (1993). The mating behaviour of tsetse flies (*Glossina*): a review. Physiological Entomology *18*, 211-218.

Le Ray, D., Barry, J.D., and Vickerman, K. (1978). Antigenic heterogeneity of metacyclic forms of *Trypanosoma brucei*. Nature *273*, 300-302.

LeBowitz, J.H., Smith, H.Q., Rusche, L., and Beverley, S.M. (1993). Coupling of poly(A) site selection and trans-splicing in *Leishmania*. Genes Dev *7*, 996-1007.

Lehane, M.J., Allingham, P.G., and Weglicki, P. (1996). Composition of the peritrophic matrix of the tsetse fly, *Glossina morsitans morsitans*. Cell Tissue Res *283*, 375-384.

Lehane, M.J., and Msangi, A.R. (1991). Lectin and peritrophic membrane development in the gut of *Glossina m.morsitans* and a discussion of their role in protecting the fly against trypanosome infection. Med Vet Entomol *5*, 495-501.

Leinhart, K., and Brown, M. (2011). SET/MYND Lysine Methyltransferases Regulate Gene Transcription and Protein Activity. Genes (Basel) *2*, 210-218.

Lenardo, M.J., Esser, K.M., Moon, A.M., Van der Ploeg, L.H., and Donelson, J.E. (1986). Metacyclic variant surface glycoprotein genes of *Trypanosoma brucei* subsp. *rhodesiense* are activated in situ, and their expression is transcriptionally regulated. Mol Cell Biol *6*, 1991-1997.

Lenardo, M.J., Rice-Ficht, A.C., Kelly, G., Esser, K.M., and Donelson, J.E. (1984). Characterization of the genes specifying two metacyclic variable antigen types in *Trypanosoma brucei rhodesiense*. Proc Natl Acad Sci U S A *81*, 6642-6646.

Liang, X.H., Haritan, A., Uliel, S., and Michaeli, S. (2003). trans and cis splicing in trypanosomatids: mechanism, factors, and regulation. Eukaryot Cell *2*, 830-840.

Lindh, J.M., Goswami, P., Blackburn, R.S., Arnold, S.E., Vale, G.A., Lehane, M.J., and Torr, S.J. (2012). Optimizing the colour and fabric of targets for the control of the tsetse fly *Glossina fuscipes fuscipes*. PLoS Negl Trop Dis *6*, e1661.

Lindner, A.K., and Priotto, G. (2010). The unknown risk of vertical transmission in sleeping sickness--a literature review. PLoS Negl Trop Dis *4*, e783.

Liu, A.Y., Van der Ploeg, L.H., Rijsewijk, F.A., and Borst, P. (1983). The transposition unit of variant surface glycoprotein gene 118 of *Trypanosoma brucei*. Presence of repeated elements at its border and absence of promoter-associated sequences. J Mol Biol *167*, 57-75.

Loveless, B.C., Mason, J.W., Sakurai, T., Inoue, N., Razavi, M., Pearson, T.W., and Boulanger, M.J. (2011). Structural characterization and epitope mapping of the glutamic acid/alanine-rich protein from *Trypanosoma congolense*: defining assembly on the parasite cell surface. J Biol Chem *286*, 20658-20665.

Lozano, R., Naghavi, M., Foreman, K., Lim, S., Shibuya, K., Aboyans, V., Abraham, J., Adair, T., Aggarwal, R., Ahn, S.Y.*, et al.* (2012). Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010. Lancet *380*, 2095-2128.

Lutumba, P., Robays, J., Miaka, C., Kande, V., Simarro, P.P., Shaw, A.P., Dujardin, B., and Boelaert, M. (2005a). The efficiency of different detection strategies of human African trypanosomiasis by *T. b. gambiense*. Trop Med Int Health *10*, 347-356.

Lutumba, P., Robays, J., Miaka mia Bilenge, C., Mesu, V.K., Molisho, D., Declercq, J., Van der Veken, W., Meheus, F., Jannin, J., and Boelaert, M. (2005b). Trypanosomiasis control, Democratic Republic of Congo, 1993-2003. Emerg Infect Dis *11*, 1382-1388.

Lyons, M. (1992). The colonial disease: a social history of sleeping sickness in northern Zaire, 1900-1940. Cambridge: Cambridge University Press.

MacGregor, P., Rojas, F., Dean, S., and Matthews, K.R. (2013). Stable transformation of pleomorphic bloodstream form *Trypanosoma brucei*. Mol Biochem Parasitol *190*, 60-62.

Magnus, E., Van Meirvenne, N., Vervoort, T., Le Ray, D., and Wery, M. (1978). Use of freeze-dried trypanosomes in the indirect fluorescent antibody test for the serodiagnosis of sleeping sickness. Ann Soc Belg Med Trop *58*, 103-109.

Malvy, D., and Chappuis, F. (2011). Sleeping sickness. Clin Microbiol Infect *17*, 986-995.

Manthri, S., Guther, M.L., Izquierdo, L., Acosta-Serrano, A., and Ferguson, M.A. (2008). Deletion of the TbALG3 gene demonstrates site-specific *N*-glycosylation and *N*-glycan processing in *Trypanosoma brucei*. Glycobiology *18*, 367-383.

Mantilla, B.S., Marchese, L., Casas-Sanchez, A., Dyer, N.A., Ejeh, N., Biran, M., Bringaud, F., Lehane, M.J., Acosta-Serrano, A., and Silber, A.M. (2017). Proline Metabolism is Essential for *Trypanosoma brucei brucei* Survival in the Tsetse Vector. PLoS Pathog *13*, e1006158.

Marcello, L., and Barry, J.D. (2007). Analysis of the VSG gene silent archive in *Trypanosoma brucei* reveals that mosaic gene expression is prominent in antigenic variation and is favored by archive substructure. Genome Res *17*, 1344-1352.

Marshall, R.D. (1972). Glycoproteins. Annu Rev Biochem *41*, 673-702.

Masocha, W., Rottenberg, M.E., and Kristensson, K. (2007). Migration of African trypanosomes across the blood-brain barrier. Physiol Behav *92*, 110-114.

Matthews, K.R. (2005). The developmental cell biology of *Trypanosoma brucei*. J Cell Sci *118*, 283-290.

Matthews, K.R., Tschudi, C., and Ullu, E. (1994). A common pyrimidine-rich motif governs trans-splicing and polyadenylation of tubulin polycistronic pre-mRNA in trypanosomes. Genes Dev *8*, 491-501.

McCann, J.R., Sheets, A.J., Grass, S., and St Geme, J.W., 3rd (2014). The *Haemophilus cryptic* genospecies Cha adhesin has at least two variants that differ in host cell binding, bacterial aggregation, and biofilm formation properties. J Bacteriol *196*, 1780-1788.

McConville, M.J., and Ferguson, M.A. (1993). The structure, biosynthesis and function of glycosylated phosphatidylinositols in the parasitic protozoa and higher eukaryotes. Biochem J *294 (Pt 2)*, 305-324.

McGuffin, L.J., Atkins, J.D., Salehe, B.R., Shuid, A.N., and Roche, D.B. (2015). IntFOLD: an integrated server for modelling protein structures and functions from amino acid sequences. Nucleic Acids Res *43*, W169-173.

Mehlert, A., Bond, C.S., and Ferguson, M.A. (2002). The glycoforms of a *Trypanosoma brucei* variant surface glycoprotein and molecular modeling of a glycosylated surface coat. Glycobiology *12*, 607-612.

Mehlert, A., Zitzmann, N., Richardson, J.M., Treumann, A., and Ferguson, M.A. (1998). The glycosylation of the variant surface glycoproteins and procyclic acidic repetitive proteins of *Trypanosoma brucei*. Mol Biochem Parasitol *91*, 145-152.

Mony, B.M., and Matthews, K.R. (2015). Assembling the components of the quorum sensing pathway in African trypanosomes. Mol Microbiol *96*, 220-232.

Mowatt, M.R., and Clayton, C.E. (1987). Developmental regulation of a novel repetitive protein of *Trypanosoma brucei*. Mol Cell Biol *7*, 2838-2844.

Mugnier, M.R., Cross, G.A., and Papavasiliou, F.N. (2015). The in vivo dynamics of antigenic variation in *Trypanosoma brucei*. Science *347*, 1470-1473.

Mugnier, M.R., Stebbins, C.E., and Papavasiliou, F.N. (2016). Masters of Disguise: Antigenic Variation and the VSG Coat in *Trypanosoma brucei*. PLoS Pathog *12*, e1005784.

Mugo, E., and Clayton, C. (2017). Expression of the RNA-binding protein RBP10 promotes the bloodstream-form differentiation state in *Trypanosoma brucei*. PLoS Pathog *13*, e1006560.

Mugo, E., Egler, F., and Clayton, C. (2017). Conversion of procyclic-form *Trypanosoma brucei* to the bloodstream form by transient expression of RBP10. Mol Biochem Parasitol *216*, 49-51.

Mumba Ngoyi, D. (2014). Performance of parasitological and molecular techniques for the diagnosis and surveillance of gambiense sleeping sickness. PLoS Negl Trop Dis *8(6): e2954*.

Mumba Ngoyi, D., Menten, J., Pyana, P.P., Buscher, P., and Lejon, V. (2013). Stage determination in sleeping sickness: comparison of two cell counting and two parasite detection techniques. Trop Med Int Health *18*, 778-782.

Munday, J.C., Eze, A.A., Baker, N., Glover, L., Clucas, C., Aguinaga Andres, D., Natto, M.J., Teka, I.A., McDonald, J., Lee, R.S*., et al.* (2014). *Trypanosoma brucei* aquaglyceroporin 2 is a high-affinity transporter for pentamidine and melaminophenyl arsenic drugs and the main genetic determinant of resistance to these drugs. J Antimicrob Chemother *69*, 651-663.

Nasab, F.P., Schulz, B.L., Gamarro, F., Parodi, A.J., and Aebi, M. (2008). All in one: *Leishmania major* STT3 proteins substitute for the whole oligosaccharyltransferase complex in *Saccharomyces cerevisiae*. Mol Biol Cell *19*, 3758-3768.

Ngomtcho, S.C.H., Weber, J.S., Ngo Bum, E., Gbem, T.T., Kelm, S., and Achukwi, M.D. (2017). Molecular screening of tsetse flies and cattle reveal different Trypanosoma species including *T. grayi* and *T. theileri* in northern Cameroon. Parasit Vectors *10*, 631.

Njiokou, F., Laveissiere, C., Simo, G., Nkinin, S., Grebaut, P., Cuny, G., and Herder, S. (2006). Wild fauna as a probable animal reservoir for *Trypanosoma brucei gambiense* in Cameroon. Infect Genet Evol *6*, 147-153.

Odiit, M., Kansiime, F., and Enyaru, J.C. (1997). Duration of symptoms and case fatality of sleeping sickness caused by *Trypanosoma brucei rhodesiense* in Tororo, Uganda. East Afr Med J *74*, 792-795.

Ooi, C.P., Smith, T.K., Gluenz, E., Wand, N.V., Vaughan, S., and Rudenko, G. (2018). Blocking variant surface glycoprotein synthesis alters endoplasmic reticulum exit sites/Golgi homeostasis in *Trypanosoma brucei*. Traffic *19*, 391-405.

Opperdoes, F.R., Baudhuin, P., Coppens, I., De Roe, C., Edwards, S.W., Weijers, P.J., and Misset, O. (1984). Purification, morphometric analysis, and characterization of the glycosomes (microbodies) of the protozoan hemoflagellate *Trypanosoma brucei*. J Cell Biol *98*, 1178-1184.

Paindavoine, P., Pays, E., Laurent, M., Geltmeyer, Y., Le Ray, D., Mehlitz, D., and Steinert, M. (1986). The use of DNA hybridization and numerical taxonomy in determining relationships between *Trypanosoma brucei* stocks and subspecies. Parasitology *92 (Pt 1)*, 31-50.

Parodi, A.J. (1993a). Biosynthesis of protein-linked oligosaccharides in trypanosomatid flagellates. Parasitol Today *9*, 373-377.

Parodi, A.J. (1993b). Protein *N*-glycosylation in trypanosomatids. A pathway with odd enzymes at both ends. Biol Res *26*, 69-75.

Parodi, A.J. (1998). The quality control of glycoprotein folding in the endoplasmic reticulum, a trip from trypanosomes to mammals. Braz J Med Biol Res *31*, 601-614.

Parodi, A.J. (2000). Protein glucosylation and its role in protein folding. Annu Rev Biochem *69*, 69-93.

Patzelt, E., Perry, K.L., and Agabian, N. (1989). Mapping of branch sites in trans-spliced pre-mRNAs of *Trypanosoma brucei*. Mol Cell Biol *9*, 4291-4297.

Peacock, L., Bailey, M., Carrington, M., and Gibson, W. (2014). Meiosis and haploid gametes in the pathogen *Trypanosoma brucei*. Curr Biol *24*, 181-186.

Pedram, M., and Donelson, J.E. (1999). The anatomy and transcription of a monocistronic expression site for a metacyclic variant surface glycoprotein gene in *Trypanosoma brucei*. J Biol Chem *274*, 16876-16883.

Peng, D., Kurup, S.P., Yao, P.Y., Minning, T.A., and Tarleton, R.L. (2014). CRISPR-Cas9-mediated single-gene and gene family disruption in *Trypanosoma cruzi*. MBio *6*, e02097-02014.

Perry, K.L., Watkins, K.P., and Agabian, N. (1987). Trypanosome mRNAs have unusual "cap 4" structures acquired by addition of a spliced leader. Proc Natl Acad Sci U S A *84*, 8190-8194.

Petersen, T.N., Brunak, S., von Heijne, G., and Nielsen, H. (2011). SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat Methods *8*, 785-786.

Picozzi, K., Fevre, E.M., Odiit, M., Carrington, M., Eisler, M.C., Maudlin, I., and Welburn, S.C. (2005). Sleeping sickness in Uganda: a thin line between two fatal diseases. BMJ *331*, 1238-1241.

Pierleoni, A., Martelli, P.L., and Casadio, R. (2008). PredGPI: a GPI-anchor predictor. BMC Bioinformatics *9*, 392.

Priotto, G., Kasparian, S., Mutombo, W., Ngouama, D., Ghorashian, S., Arnold, U., Ghabri, S., Baudin, E., Buard, V., Kazadi-Kyanza, S.*, et al.* (2009). Nifurtimox-eflornithine combination therapy for second-stage African *Trypanosoma brucei gambiense* trypanosomiasis: a multicentre, randomised, phase III, non-inferiority trial. Lancet *374*, 56-64.

Reuner, B., Vassella, E., Yutzy, B., and Boshart, M. (1997). Cell density triggers slender to stumpy differentiation of *Trypanosoma brucei* bloodstream forms in culture. Mol Biochem Parasitol *90*, 269-280.

Rico, E., Jeacock, L., Kovarova, J., and Horn, D. (2018). Inducible high-efficiency CRISPR-Cas9-targeted gene editing and precision base editing in African trypanosomes. Sci Rep *8*, 7960.

Rijo-Ferreira, F., Carvalho, T., Afonso, C., Sanches-Vaz, M., Costa, R.M., Figueiredo, L.M., and Takahashi, J.S. (2018). Sleeping sickness is a circadian disorder. Nat Commun *9*, 62.

Rijo-Ferreira, F., Pinto-Neves, D., Barbosa-Morais, N.L., Takahashi, J.S., and Figueiredo, L.M. (2017a). *Trypanosoma brucei* metabolism is under circadian control. Nat Microbiol *2*, 17032.

Rijo-Ferreira, F., Takahashi, J.S., and Figueiredo, L.M. (2017b). Circadian rhythms in parasites. PLoS Pathog *13*, e1006590.

Robays, J., Bilengue, M.M., Van der Stuyft, P., and Boelaert, M. (2004). The effectiveness of active population screening and treatment for sleeping sickness control in the Democratic Republic of Congo. Trop Med Int Health *9*, 542-550.

Rocha, G., Martins, A., Gama, G., Brandao, F., and Atouguia, J. (2004). Possible cases of sexual and congenital transmission of sleeping sickness. Lancet *363*, 247.

Roditi, I., Carrington, M., and Turner, M. (1987). Expression of a polypeptide containing a dipeptide repeat is confined to the insect stage of *Trypanosoma brucei*. Nature *325*, 272-274.

Roditi, I., and Clayton, C. (1999). An unambiguous nomenclature for the major surface glycoproteins of the procyclic form of *Trypanosoma brucei*. Mol Biochem Parasitol *103*, 99-100.

Roditi, I., Schwarz, H., Pearson, T.W., Beecroft, R.P., Liu, M.K., Richardson, J.P., Buhring, H.J., Pleiss, J., Bulow, R., Williams, R.O.*, et al.* (1989). Procyclin gene expression and loss of the variant surface glycoprotein during differentiation of *Trypanosoma brucei*. J Cell Biol *108*, 737-746.

Rogers, M.E. (2012). The role of *Leishmania* proteophosphoglycans in sand fly transmission and infection of the Mammalian host. Front Microbiol *3*, 223.

Rose, C., Belmonte, R., Armstrong, S.D., Molyneux, G., Haines, L.R., Lehane, M.J., Wastling, J., and Acosta-Serrano, A. (2014). An investigation into the protein composition of the teneral *Glossina morsitans morsitans* peritrophic matrix. PLoS Negl Trop Dis *8*, e2691.

Rotureau, B., Subota, I., Buisson, J., and Bastin, P. (2012). A new asymmetric division contributes to the continuous production of infective trypanosomes in the tsetse fly. Development *139*, 1842-1850.

Roy, A., Kucukural, A., and Zhang, Y. (2010). I-TASSER: a unified platform for automated protein structure and function prediction. Nat Protoc *5*, 725-738.

Sakurai, T., Sugimoto, C., and Inoue, N. (2008). Identification and molecular characterization of a novel stage-specific surface protein of *Trypanosoma congolense* epimastigotes. Mol Biochem Parasitol *161*, 1-11.

Samuelson, J., Banerjee, S., Magnelli, P., Cui, J., Kelleher, D.J., Gilmore, R., and Robbins, P.W. (2005). The diversity of dolichol-linked precursors to Asn-linked glycans likely results from secondary loss of sets of glycosyltransferases. Proc Natl Acad Sci U S A *102*, 1548-1553.

Sanz, S., Lopez-Gutierrez, B., Bandini, G., Damerow, S., Absalon, S., Dinglasan, R.R., Samuelson, J., and Izquierdo, L. (2016). The disruption of GDP-fucose de novo biosynthesis suggests the presence of a novel fucose-containing glycoconjugate in *Plasmodium* asexual blood stages. Sci Rep *6*, 37230.

Savage, A.F., Cerqueira, G.C., Regmi, S., Wu, Y., El Sayed, N.M., and Aksoy, S. (2012). Transcript expression analysis of putative *Trypanosoma brucei* GPI-anchored surface proteins during development in the tsetse and mammalian hosts. PLoS Negl Trop Dis *6*, e1708.

Savage, A.F., Kolev, N.G., Franklin, J.B., Vigneron, A., Aksoy, S., and Tschudi, C. (2016). Transcriptome Profiling of *Trypanosoma brucei* Development in the Tsetse Fly Vector *Glossina morsitans*. PLoS One *11*, e0168877.

Schmid, C., Kuemmerle, A., Blum, J., Ghabri, S., Kande, V., Mutombo, W., Ilunga, M., Lumpungu, I., Mutanda, S., Nganzobo, P.*, et al.* (2012). In-hospital safety in field conditions of nifurtimox eflornithine combination therapy (NECT) for *T. b. gambiense* sleeping sickness. PLoS Negl Trop Dis *6*, e1920.

Schmid, C., Richer, M., Bilenge, C.M., Josenando, T., Chappuis, F., Manthelot, C.R., Nangouma, A., Doua, F., Asumu, P.N., Simarro, P.P.*, et al.* (2005). Effectiveness of a 10-day melarsoprol schedule for the treatment of late-stage human African trypanosomiasis: confirmation from a multinational study (IMPAMEL II). J Infect Dis *191*, 1922-1931.

Schumann Burkard, G., Kaser, S., de Araujo, P.R., Schimanski, B., Naguleswaran, A., Knusel, S., Heller, M., and Roditi, I. (2013). Nucleolar proteins regulate stage-

specific gene expression and ribosomal RNA maturation in *Trypanosoma brucei*. Mol Microbiol *88*, 827-840.

Schuster, S., Kruger, T., Subota, I., Thusek, S., Rotureau, B., Beilhack, A., and Engstler, M. (2017). Developmental adaptations of trypanosome motility to the tsetse fly host environments unravel a multifaceted in vivo microswimmer system. Elife *6*.

Schwede, A., Jones, N., Engstler, M., and Carrington, M. (2011). The VSG C-terminal domain is inaccessible to antibodies on live trypanosomes. Mol Biochem Parasitol *175*, 201-204.

Schwede, A., Macleod, O.J., MacGregor, P., and Carrington, M. (2015). How Does the VSG Coat of Bloodstream Form African Trypanosomes Interact with External Proteins? PLoS Pathog *11*, e1005259.

Serricchio, M., and Butikofer, P. (2013). Phosphatidylglycerophosphate synthase associates with a mitochondrial inner membrane complex and is essential for growth of *Trypanosoma brucei*. Mol Microbiol *87*, 569-579.

Shaner, N.C., Lambert, G.G., Chammas, A., Ni, Y., Cranfill, P.J., Baird, M.A., Sell, B.R., Allen, J.R., Day, R.N., Israelsson, M.*, et al.* (2013). A bright monomeric green fluorescent protein derived from *Branchiostoma lanceolatum*. Nat Methods *10*, 407-409.

Sharma, R., Gluenz, E., Peacock, L., Gibson, W., Gull, K., and Carrington, M. (2009). The heart of darkness: growth and form of *Trypanosoma brucei* in the tsetse fly. Trends Parasitol *25*, 517-524.

Sharma, R., Peacock, L., Gluenz, E., Gull, K., Gibson, W., and Carrington, M. (2008). Asymmetric cell division as a route to reduction in cell length and change in cell morphology in trypanosomes. Protist *159*, 137-151.

Shaw, A.P., Torr, S.J., Waiswa, C., Cecchi, G., Wint, G.R., Mattioli, R.C., and Robinson, T.P. (2013). Estimating the costs of tsetse control options: an example for Uganda. Prev Vet Med *110*, 290-303.

Shaw, A.P., Wint, G.R., Cecchi, G., Torr, S.J., Mattioli, R.C., and Robinson, T.P. (2015). Mapping the benefit-cost ratios of interventions against bovine trypanosomosis in Eastern Africa. Prev Vet Med *122*, 406-416.

Shereni, W., Anderson, N.E., Nyakupinda, L., and Cecchi, G. (2016). Spatial distribution and trypanosome infection of tsetse flies in the sleeping sickness focus of Zimbabwe in Hurungwe District. Parasit Vectors *9*, 605.

Siegel, T.N., Gunasekera, K., Cross, G.A., and Ochsenreiter, T. (2011). Gene expression in *Trypanosoma brucei*: lessons from high-throughput RNA sequencing. Trends Parasitol *27*, 434-441.

Sienkiewicz, N., Jaroslawski, S., Wyllie, S., and Fairlamb, A.H. (2008). Chemical and genetic validation of dihydrofolate reductase-thymidylate synthase as a drug target in African trypanosomes. Mol Microbiol *69*, 520-533.

Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Soding, J.*, et al.* (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol *7*, 539.

Simarro, P.P., Cecchi, G., Franco, J.R., Paone, M., Diarra, A., Priotto, G., Mattioli, R.C., and Jannin, J.G. (2015). Monitoring the Progress towards the Elimination of Gambiense Human African Trypanosomiasis. PLoS Negl Trop Dis *9*, e0003785.

Simarro, P.P., Cecchi, G., Franco, J.R., Paone, M., Diarra, A., Ruiz-Postigo, J.A., Fevre, E.M., Mattioli, R.C., and Jannin, J.G. (2012). Estimating and mapping the population at risk of sleeping sickness. PLoS Negl Trop Dis *6*, e1859.

Simarro, P.P., Cecchi, G., Paone, M., Franco, J.R., Diarra, A., Ruiz, J.A., Fevre, E.M., Courtin, F., Mattioli, R.C., and Jannin, J.G. (2010). The Atlas of human African trypanosomiasis: a contribution to global mapping of neglected tropical diseases. Int J Health Geogr *9*, 57.

Simo, G., Asonganyi, T., Nkinin, S.W., Njiokou, F., and Herder, S. (2006). High prevalence of *Trypanosoma brucei gambiense* group 1 in pigs from the Fontem sleeping sickness focus in Cameroon. Vet Parasitol *139*, 57-66.

Simpson, A.M., and Simpson, L. (1980). Kinetoplast DNA and RNA of *Trypanosoma brucei*. Mol Biochem Parasitol *2*, 93-108.

Sistrom, M., Evans, B., Benoit, J., Balmer, O., Aksoy, S., and Caccone, A. (2016). De Novo Genome Assembly Shows Genome Wide Similarity between *Trypanosoma brucei brucei* and *Trypanosoma brucei rhodesiense*. PLoS One *11*, e0147660.

Smith, D.H., Pepin, J., and Stich, A.H. (1998). Human African trypanosomiasis: an emerging public health crisis. Br Med Bull *54*, 341-355.

Smith, H.Q., Li, C., Stanley, C.A., and Smith, T.J. (2017). Glutamate Dehydrogenase, a Complex Enzyme at a Crucial Metabolic Branch Point. Neurochem Res.

Sollelis, L., Ghorbal, M., MacPherson, C.R., Martins, R.M., Kuk, N., Crobu, L., Bastien, P., Scherf, A., Lopez-Rubio, J.J., and Sterkers, Y. (2015). First efficient CRISPR-Cas9-mediated genome editing in *Leishmania* parasites. Cell Microbiol *17*, 1405-1412.

Son, H.F., Kim, I.K., and Kim, K.J. (2015). Structural insights into domain movement and cofactor specificity of glutamate dehydrogenase from *Corynebacterium glutamicum*. Biochem Biophys Res Commun *459*, 387-392.

Song, J., Tan, H., Perry, A.J., Akutsu, T., Webb, G.I., Whisstock, J.C., and Pike, R.N. (2012). PROSPER: an integrated feature-based tool for predicting protease substrate cleavage sites. PLoS One *7*, e50300.

Spellmon, N., Holcomb, J., Trescott, L., Sirinupong, N., and Yang, Z. (2015). Structure and function of SET and MYND domain-containing proteins. Int J Mol Sci *16*, 1406-1428.

Sterkel, M., Perdomo, H.D., Guizzo, M.G., Barletta, A.B., Nunes, R.D., Dias, F.A., Sorgine, M.H., and Oliveira, P.L. (2016). Tyrosine Detoxification Is an Essential Trait in the Life History of Blood-Feeding Arthropods. Curr Biol *26*, 2188-2193.

Stevens, J. (2004). Systematics of trypanosomes of medical and veterinary importance. Wallingford: CABI Publishing *1-23*.

Stodkilde, K., Torvund-Jensen, M., Moestrup, S.K., and Andersen, C.B. (2014). Structural basis for trypanosomal haem acquisition and susceptibility to the host innate immune system. Nat Commun *5*, 5487.

Szempruch, A.J., Sykes, S.E., Kieft, R., Dennison, L., Becker, A.C., Gartrell, A., Martin, W.J., Nakayasu, E.S., Almeida, I.C., Hajduk, S.L*., et al.* (2016). Extracellular Vesicles from *Trypanosoma brucei* Mediate Virulence Factor Transfer and Cause Host Anemia. Cell *164*, 246-257.

Takatsuki, A., Arima, K., and Tamura, G. (1971). Tunicamycin, a new antibiotic. I. Isolation and characterization of tunicamycin. J Antibiot (Tokyo) *24*, 215-223.

Tarral, A., Blesson, S., Mordt, O.V., Torreele, E., Sassella, D., Bray, M.A., Hovsepian, L., Evene, E., Gualano, V., Felices, M*., et al.* (2014). Determination of an optimal dosing regimen for fexinidazole, a novel oral drug for the treatment of human African trypanosomiasis: first-in-human studies. Clin Pharmacokinet *53*, 565-580.

Telleria, E.L., Benoit, J.B., Zhao, X., Savage, A.F., Regmi, S., Alves e Silva, T.L., O'Neill, M., and Aksoy, S. (2014). Insights into the trypanosome-host interactions revealed through transcriptomic analysis of parasitized tsetse fly salivary glands. PLoS Negl Trop Dis *8*, e2649.

Tetley, L., Turner, C.M., Barry, J.D., Crowe, J.S., and Vickerman, K. (1987). Onset of expression of the variant surface glycoproteins of *Trypanosoma brucei* in the tsetse fly studied using immunoelectron microscopy. J Cell Sci *87 (Pt 2)*, 363-372.

Thomson, R., Genovese, G., Canon, C., Kovacsics, D., Higgins, M.K., Carrington, M., Winkler, C.A., Kopp, J., Rotimi, C., Adeyemo, A*., et al.* (2014). Evolution of the primate trypanolytic factor APOL1. Proc Natl Acad Sci U S A *111*, E2130-2139.

Tiengwe, C., Bush, P.J., and Bangs, J.D. (2017). Controlling transferrin receptor trafficking with GPI-valence in bloodstream stage African trypanosomes. PLoS Pathog *13*, e1006366.

Torr, S.J., Maudlin, I., and Vale, G.A. (2007). Less is more: restricted application of insecticide to cattle to improve the cost and efficacy of tsetse control. Med Vet Entomol *21*, 53-64.

Trindade, S., Rijo-Ferreira, F., Carvalho, T., Pinto-Neves, D., Guegan, F., Aresta-Branco, F., Bento, F., Young, S.A., Pinto, A., Van Den Abbeele, J*., et al.* (2016). *Trypanosoma brucei* Parasites Occupy and Functionally Adapt to the Adipose Tissue in Mice. Cell Host Microbe *19*, 837-848.

Turner, C.M., Barry, J.D., Maudlin, I., and Vickerman, K. (1988). An estimate of the size of the metacyclic variable antigen repertoire of *Trypanosoma brucei rhodesiense*. Parasitology *97 (Pt 2)*, 269-276.

Urwyler, S., Studer, E., Renggli, C.K., and Roditi, I. (2007). A family of stage-specific alanine-rich proteins on the surface of epimastigote forms of *Trypanosoma brucei*. Mol Microbiol *63*, 218-228.

Urwyler, S., Vassella, E., Van Den Abbeele, J., Renggli, C.K., Blundell, P., Barry, J.D., and Roditi, I. (2005). Expression of procyclin mRNAs during cyclical transmission of *Trypanosoma brucei*. PLoS Pathog *1*, e22.

Uzureau, P., Uzureau, S., Lecordier, L., Fontaine, F., Tebabi, P., Homble, F., Grelard, A., Zhendre, V., Nolan, D.P., Lins, L.*, et al.* (2013). Mechanism of *Trypanosoma brucei gambiense* resistance to human serum. Nature *501*, 430-434.

Van Den Abbeele, J., Caljon, G., De Ridder, K., De Baetselier, P., and Coosemans, M. (2010). *Trypanosoma brucei* modifies the tsetse salivary composition, altering the fly feeding behavior that favors parasite transmission. PLoS Pathog *6*, e1000926.

Van Den Abbeele, J., Claes, Y., van Bockstaele, D., Le Ray, D., and Coosemans, M. (1999). *Trypanosoma brucei* spp. development in the tsetse fly: characterization of the post-mesocyclic stages in the foregut and proboscis. Parasitology *118 (Pt 5)*, 469-478.

Van Hoof, L., Henrard, C., and Peel, E. (1937). Influences modificatricees de la transmissibilité cylique du *Trypanosoma gambiense* par *Glossina palpalis*. Ann Soc Belg Med Trop *17*, 249-263.

Van Nieuwenhove, S., Betu-Ku-Mesu, V.K., Diabakana, P.M., Declercq, J., and Bilenge, C.M. (2001). Sleeping sickness resurgence in the DRC: the past decade. Trop Med Int Health *6*, 335-341.

van Nieuwenhoven, M.A., Kovacs, E.M., Brummer, R.J., Westerterp-Plantenga, M.S., and Brouns, F. (2001). The effect of different dosages of guar gum on gastric emptying and small intestinal transit of a consumed semisolid meal. J Am Coll Nutr *20*, 87-91.

Vanhamme, L., and Pays, E. (2004). The trypanosome lytic factor of human serum and the molecular basis of sleeping sickness. Int J Parasitol *34*, 887-898.

Varki, A. (1993). Biological roles of oligosaccharides: all of the theories are correct. Glycobiology *3*, 97-130.

Varki, A., Schnaar, R.L., and Schauer, R. (2015). N-glycosylation. In Essentials of Glycobiology, rd, A. Varki, R.D. Cummings, J.D. Esko, P. Stanley, G.W. Hart, M. Aebi, A.G. Darvill, T. Kinoshita, N.H. Packer*, et al.*, eds. (Cold Spring Harbor (NY)), pp. 179-195.

Vasquez, J.J., Hon, C.C., Vanselow, J.T., Schlosser, A., and Siegel, T.N. (2014). Comparative ribosome profiling reveals extensive translational complexity in different *Trypanosoma brucei* life cycle stages. Nucleic Acids Res *42*, 3623-3637.

Vasquez, J.J., Wedel, C., Cosentino, R.O., and Siegel, T.N. (2018). Exploiting CRISPR-Cas9 technology to investigate individual histone modifications. Nucleic Acids Res.

Vassella, E., Den Abbeele, J.V., Butikofer, P., Renggli, C.K., Furger, A., Brun, R., and Roditi, I. (2000). A major surface glycoprotein of *Trypanosoma brucei* is expressed transiently during development and can be regulated post-transcriptionally by glycerol or hypoxia. Genes Dev *14*, 615-626.

Vassella, E., Oberle, M., Urwyler, S., Renggli, C.K., Studer, E., Hemphill, A., Fragoso, C., Butikofer, P., Brun, R., and Roditi, I. (2009). Major surface glycoproteins of insect forms of *Trypanosoma brucei* are not essential for cyclical transmission by tsetse. PLoS One *4*, e4493.

Vergult, S., Hoogeboom, A.J., Bijlsma, E.K., Sante, T., Klopocki, E., De Wilde, B., Jongmans, M., Thiel, C., Verheij, J.B., Perez-Aytes, A.*, et al.* (2013). Complex genetics of radial ray deficiencies: screening of a cohort of 54 patients. Genet Med *15*, 195-202.

Vickerman, K. (1985). Developmental cycles and biology of pathogenic trypanosomes. Br Med Bull *41*, 105-114.

Wang, J., Weiss, B.L., and Aksoy, S. (2013). Tsetse fly microbiota: form and function. Front Cell Infect Microbiol *3*, 69.

Waterhouse, A.M., Procter, J.B., Martin, D.M., Clamp, M., and Barton, G.J. (2009). Jalview Version 2--a multiple sequence alignment editor and analysis workbench. Bioinformatics *25*, 1189-1191.

Webb, B., Sali, A. (2016). Protein Structure Modelling with MODELLER. Functional Genomics, 39-54.

Webster, P., and Russell, D.G. (1993). The flagellar pocket of trypanosomatids. Parasitol Today *9*, 201-206.

Weiss, B.L., Wang, J., Maltz, M.A., Wu, Y., and Aksoy, S. (2013). Trypanosome infection establishment in the tsetse fly gut is influenced by microbiome-regulated host immune barriers. PLoS Pathog *9*, e1003318.

Welburn, S.C., Picozzi, K., Fevre, E.M., Coleman, P.G., Odiit, M., Carrington, M., and Maudlin, I. (2001). Identification of human-infective trypanosomes in animal reservoir of sleeping sickness in Uganda by means of serum-resistance-associated (SRA) gene. Lancet *358*, 2017-2019.

Wendo, C. (2002). Uganda revises cattle treatment to protect humans from sleeping sickness. Lancet *359*, 239.

Wigglesworth, V.B. (1929). Digestion in the tsetse-fly: a study of structure and function. Parasitology *21*.

Wijers, D.J. (1958). Factors that may influence the infection rate of *Glossina palpalis* with *Trypanosoma gambiense*. I. The age of the fly at the time of the infected feed. Ann Trop Med Parasitol *52*, 385-390.

Williams, R.O., Young, J.R., and Majiwa, P.A. (1982). Genomic environment of *T. brucei* VSG genes: presence of a minichromosome. Nature *299*, 417-421.

Wirtz, E., Leal, S., Ochatt, C., and Cross, G.A. (1999). A tightly regulated inducible expression system for conditional gene knock-outs and dominant-negative genetics in *Trypanosoma brucei*. Mol Biochem Parasitol *99*, 89-101.

Woo, P.T. (1970). The haematocrit centrifuge technique for the diagnosis of African trypanosomiasis. Acta Trop *27*, 384-386.

Yang, J., Yan, R., Roy, A., Xu, D., Poisson, J., and Zhang, Y. (2015). The I-TASSER Suite: protein structure and function prediction. Nat Methods *12*, 7-8.

Zamze, S.E., Ferguson, M.A., Collins, R., Dwek, R.A., and Rademacher, T.W. (1988). Characterization of the cross-reacting determinant (CRD) of the glycosyl-phosphatidylinositol membrane anchor of *Trypanosoma brucei* variant surface glycoprotein. Eur J Biochem *176*, 527-534.

Zhang, Y. (2008). I-TASSER server for protein 3D structure prediction. BMC Bioinformatics *9*, 40.

Ziegelbauer, K., Quinten, M., Schwarz, H., Pearson, T.W., and Overath, P. (1990). Synchronous differentiation of *Trypanosoma brucei* from bloodstream to procyclic forms in vitro. Eur J Biochem *192*, 373-378.

Zimmermann, H., Subota, I., Batram, C., Kramer, S., Janzen, C.J., Jones, N.G., and Engstler, M. (2017). A quorum sensing-independent path to stumpy development in *Trypanosoma brucei*. PLoS Pathog *13*, e1006324.

Zoll, S., Lane-Serff, H., Mehmood, S., Schneider, J., Robinson, C.V., Carrington, M., and Higgins, M.K. (2018). The structure of serum resistance-associated protein and its implications for human African trypanosomiasis. Nat Microbiol *3*, 295-301.

Zomerdijk, J.C., Ouellette, M., ten Asbroek, A.L., Kieft, R., Bommer, A.M., Clayton, C.E., and Borst, P. (1990). The promoter for a variant surface glycoprotein gene expression site in *Trypanosoma brucei*. EMBO J *9*, 2791-2801.

APPENDIX I

Primer Code	Sequence 5' -> 3'
3xHA_sfGFP_F_HindIII	CGCAAGCTTATGTACCATATGATGTGCCTGAC
3xHA_sfGFP_R_XhoI	CGCCTCGAGCTATACCATATGATGTGCCCTGAC
BARP_qPCR_F	AAGCAAAGGTACAAGCAGAG
BARP_qPCR_R	CGAGTGTTGCTCTCACAG
Cvar_F	GCCGAAAGGACGGCAGAGACG
Cvar_R	GTGTATCCTCCTATAGATTCTGCATAGC
FT_CDS_R_BamHI	CGCGGATCCTCAGCGTCTGTCGCGTCTGCCGT
FT_HA_CDS_F_HindIII	CGCAAGCTTATG-[HA]-GCTTCGACAATCTACAATTTCAA
GDH_CDS_F_XbaI	CGCTCTAGAATGATGCGCCGTGCTTCTTA
GDH_CDS_R_XbaI	CGCTCTAGAAGAGATGTGCTTGGCCTGC
MISP_-GPI_R_XbaI	CGCTCTAGAGCTGCCGTCATTACCAAC
MISP_-GPI_R_XhoI	GCGCTCGAGGCTGCCGTCATTACCAAC
MISP_-SP_F_NcoI	GCGCCATGGACTCCATAATTGAGGAAGGG
MISP_-SP_F_NdeI	CGCCATATGGACTCCATAATTGAGGAAGGG
MISP_C_F_NdeI	CGCCATATGACGGAAGAAGCGGATGTTC
MISP_qPCR_F	GTGAGGAAGCAGAAGTTGG
MISP_qPCR_R	AAAGTGCTGCAAGAAGGATC
MISP_RNAi_ANTI_F_NdeI	CGCCATATGCACTACAGCAATTCGGGATAATC
MISP_RNAi_ANTI_R_HindIII	GCGAAGCTTGCCTCTCGCAATTCGTCTTG
MISP_RNAi_SENSE_F_XhoI	CGCCTCGAGCACTACAGCAATTCGGGATAATC
MISP_RNAi_SENSE_R_BamHI	GCGGGATCCGCCTCTCGCAATTCGTCTTG
MISP360_CDS_R_BglII	GCGAGATCTTTAAAAATGTGCGGCAGC
MISP360_CDS-SP_F_XbaI	GCGTCTAGAGACTCCATAATTGAGGAAGG
mNG_CDS_F_HindIII	CGCAAGCTTATGGTCTCGAAAGGTGAGGAAG
mNG_CDS_R_BamHI	CGCGGATCCCTACTTATACAATCGTCCATCCC
MYND_CDS_F_HindIII	CGCAAGCTTATGACAACTTCCCCATCCAC
MYND_CDS_R_XbaI	CGCTCTAGAGCTGAAGGCAGCGGACTTT
RBP6_CDS_R_BamHI	CGCGGATCCTCAACCAGCGGCACCGCGGGAAC
RBP6_HA_CDS_F_HindIII	CGCAAGCTTATG-[HA]-CTTTCTACCCCAACAGCCCGCA
RBP6_RNAi_ANTI_F_NdeI	CGCCATATGCAACAGCCGTATCATCCCTT
RBP6_RNAi_ANTI_R_HindIII	CGCAAGCTTACTCATTCGCTCCACAGCTT
RBP6_RNAi_SENSE_F_XhoI	CGCCTCGAGCAACAGCCGTATCATCCCTT

RBP6_RNAi_SENSE_R_BamHI	CGCGGATCCACTCATTCGCTCCACAGCTT
RFT1_3_F_XbaI	TCTAAGGTGGGAATAATG
RFT1_3_R_SacI	AGCTTGGAGTCCATGAC
RFT1_5_F_HindIII	GGAAGCGCAATCATTCAGAG
RFT1_5_R_EcoRI	CACACCAAAGGTACAGCTGC
RFT1_CDS_F_HindIII	ATGGACTTCAAACGACAGCTGGC
RFT1_CDS_R_XhoI	CTACTCGCCCGTTG
SST3B_RT_F	GCGTCGAATGATGATTTCTTTG
SST3C_RT_F	GGCCTTCCTACGTCATC
SST3C_RT_R	TCATCGGCAAGAACCAAC
STT3_3_F_XbaI	CGCTCTAGAACCCATGTTTCCGTCTTG
STT3_3_R_SacI	CGCGAGCTCCCCGCTATAAGCTACTTCTC
STT3_5_F_HindIII	CGCAAGCTTGCGTATGTGCCTTGTTAGTG
STT3_5_R_EcoRI	CGCGAATTCTTATGAAATGGGGATGAG
STT3_RNAi_ANTI_F_NdeI	CGCCATATGGTGAGGGAGGGATGAGTAAG
STT3_RNAi_ANTI_R_HindIII	CGCAAGCTTCCATTGATCGCATCAGGTG
STT3_RNAi_SENSE_F_XhoI	CGCCTCGAGGTGAGGAGGGATGAGTAAG
STT3_RNAi_SENSE_R_BamHI	CGCGGATCCCCCATTGATCGCATCAGGTG
STT3A_RT_F	ATTGTGGCAAGTTTGCTTG
STT3A_RT_R	CATTGGGTAGATCAGTCACG
STT3B_RT_R	TATAATCTGGGGGCTCGAC
TERT_qPCR_F	AACAGTAAGTCCCGTGTGG
TERT_qPCR_R	GCCTTCAGTTTGTCCAAGAAG
TERT_RT-PCR_F	AACAGTAAGTCCCGTGTGG
TERT_RT-PCR_R	CAAAGTCGTACTTCAACATCAG